AD-A262 578

(2)

RL-TR-92-273
Final Technical Report
December 1992

NATURAL LANGUAGE GENERATION

20001013205

USC Information Sciences Institute

Eduard H. Hovy

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-92-273 has been reviewed and is approved for publication.

APPROVED:

DOUGLAS A. WHITE
Project Engineer

FOR THE COMMANDER:

JOHN A. GRANIERO
Chief Scientist
Command, Control and Communications Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL(C3CA) Griffiss AFB, NY 13441-4505. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1992	Final Mar 89 - Dec 91

4. TITLE AND SUBTITLE
NATURAL LANGUACE CENERATION

5. FUNDING NUMBERS
C - F30602-89-C-0041
PE - 62702F
PR - 5581
TA - 27
WU - 47

6. AUTHOR(S)
Eduard H. Hovy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USC Information Sciences Institute
4676 Admiralty Way
Marina Del Rey CA 90292-6695

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (C3CA)
525 Brooks Road
Griffiss AFB NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-92-273

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Dougals A. White/C3CA/(315)330-3564

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This document reports on research investigating the automatic planning and generation of multisentence text and multimodal presentations by computer. This work was performed from mid-1989 to late 1991 at the Information Services Institute of the University of Southern California (USC/ISI). The research consisted of three stages: development of the basic planning algorithms, collection and synthesis of subsidiary information required for text and multimedia presentations, and design and construction of a new type of communication planner architecture.

14. SUBJECT TERMS
Text generation, natural language, artificial intelligence, computer generated test.

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

FINAL REPORT: NATURAL LANGUAGE GENERATION

Eduard H. Hovy

Information Sciences Institute
of the University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Telephone: 310-822-1511
Fax: 310-823-6714
Email: HOVY@ISI.EDU

December 1991

Final Report of work done under
RADC Contract No. F30602-89-C-0041

Contents

1 Introduction

This document reports on the research done over a period of approximately two and a half years (from mid-1989 to end 1991) investigating the automatic planning and generation of multisentence text by computer, at the Information Sciences Institute of the University of Southern California (USC/ISI), under funding from the Rome Laboratories of the U.S. Air Force.

The research can be broken into three stages. During the first stage, which lasted for about ten months, the basic text planning paradigm developed in 1988-89 at USC/ISI was thoroughly investigated. The satisfactoriness of a top-down stepwise refinement procedure, using interclausal relations from Rhetorical Structure Theory (RST) [Mann & Thompson 88] as plan operators, was demonstrated. The need for controlling planning using additional linguistic constraints (such as focus shift) was explored in a preliminary experiment. The need for work to be performed in several ancillary areas of text planning, such as the development of a satisfactorily encompassing library of intersegment relations, the need for a powerful notation with which to represent speaker intentionality, and the need for a powerful theory of sentence-level planning after text structuring, was demonstrated. Solutions in all or most of these areas were necessary before a powerful enough text planner could be developed to produce page-length text in specific domains.

The second stage, which lasted for about a year, involved the collection and synthesis of information regarding some of these areas. An extensive survey of several hundred proposed discourse structure relations was executed and the results were taxonomized into a hierarchy of relations organized on functional principles. The applicability of the text planning structures and techniques was demonstrated for automated text formatting. Initial investigations were conducted into the applicability of some of the same representational and procedural techniques used for text planning on the problem of automated multimedia planning.

The third stage involved the design and construction of a totally new type of text planner architecture as required to handle the complexity of the disparate types of knowledge that play a role in determining text structure and content. Work on the multimedia aspects of planning human-computer communications was continued and refined.

Throughout this time, efforts were made to broadcast the problems and strengths of this work both within the U.S. and internationally, in order to accelerate the development of these ideas into the maturity of a well-tested foundation which would support the construction of general-purpose multisentence text planners to complement the recent first appearance of "general-purpose" single-sentence generators such as Penman, MUMBLE, and FUF.

This document first describes the technical work and then briefly outlines the efforts of outreach. The technical work is described from the point of view of an emerging theory of discourse — multimodal human-computer interactive discourse — along the following lines: First, the basic

problem of discourse and text planning is described. Next the first text planning experiments are outlined in some detail. The issues that resulted from those experiments, and the work done on them, follows. The next section describes the design of a new type of text planner architecture. The last two technical sections are devoted to purely multimedia questions. The final section of this document outlines the outreach and dissemination of the ideas developed under the contract.

2 Objectives

Given the complexity of building autonomous non-human intelligent agents, it has become clear that for a considerable time, if not for ever, humans and computers are going to be performing tasks as cooperating partners. This development means that a great deal of effort must be placed on developing powerful, efficient, and natural ways of communicating between people and computers. Since it is proving feasible (though not easy) to analyze and generate human language into and from computer-internal format in restricted domains, and since the cost of teaching people specialized computer languages and interaction procedures is likely to remain high, it is incumbent on Artificial Intelligence researchers to develop algorithms that support human-computer interactions of the most powerful kind: using human language and additional media, as natural and appropriate.

The work presented in this report is a step toward meeting some of the most critical needs in human-computer communication on various fronts: in multisentence natural language generation (NLG), in the development of notations that support the understanding of discourse, and in the development of theory and techniques to support the multimodal display of information.

Unfortunately, despite the increasing needs for concise and clear output from computers which contain ever-increasing amounts of data and perform at ever-higher speeds, NL generation technology has not enjoyed a great deal of research support when compared with NL analysis. It is hampered by a short technical history, the (incorrect) general belief that generation is "easier" than parsing, the complexity of controlling language behavior on pragmatic and non-linguistic grounds, and relatively little understanding of how language works at larger-than-sentence levels. The research and development performed under this contract concentrated on the last point, namely the development of general domain-independent techniques for planning coherent multisentence paragraphs of text. These techniques are integrated with well-established single-sentence generation technology and made suitable for effective inclusion in an integrated multimodal interface environment.

Under separate funding, additional research is being performed at USC/ISI on the development of grammars and semantic representations within the context of machine translation: that is, on representations that support both analysis of language and generation in various languages. In addition, other projects are actively involved in using language generation for explaining expert

system behavior and for generating descriptions of software under development. This work provides a rich context in which the work described here took place.

3 Technical Work

3.1 The Problem: Discourse and Text Planning

Every day, we effortlessly produce thousands of words of connected discourse from complicated and ill-understood internal knowledge for complicated and ill-understood reasons. In spite of over three decades of work on natural language processing, computers are nowhere near this capability. However, computational efforts to mimic the generation processes have, over the past decade, established the power of viewing language generation as a goal-driven and hence essentially planning process (in contrast to analysis, which is input-driven and essentially a process of inference). This perspective mandates the formulation of plans and planners that govern the selection and assembly of material into coherent grammatical text in order to achieve the author's communicative purposes.

In this document we focus on discourse structure as seen from the planning perspective of generation. We argue that without such a notion, communication is unlikely to succeed. We outline various theories of discourse structure, linguistic and computational. We describe a series of computational experiments, conducted at various locations, that investigated several of the major problems that arise when one tries to plan discourse automatically, and show the central role played by discourse relations in making up and giving form to discourse.

As an initial assumption, we take it that discourse is a goal-oriented phenomenon: people communicate for a reason. Though these goals do not always decompose into a structure of increasingly specific subgoals — think of interacting with a 4-year-old, joking in a supermarket line, reminiscing around a fire — enough of them do to make the traditional Artificial Intelligence planning approach (goal decomposition) rewarding. Discourses that do admit such an analysis are typically informative messages such as annual reports and encyclopedia entries, instructions, explanations, and other collaborations toward some purpose — the kinds of things we want computers to do in any case.

We discuss only monologic discourse here; the additional issues that are required for multi-party discourse are still at early stages of study.

3.1.1 Discourse Structure

Computational research on understanding and producing language has concentrated largely on single-sentence phenomena. Though today there are numerous parsers, semantic analyzers, sentence

generators, lexicon acquisition tools, etc., available, not more than a handful of systems claim to perform multisentence analysis or generation on more than a toy scale.

Of course, no account of language that stops at the sentence level is adequate, and neither are programs that communicate solely on the sentence level. But moving "up" from the sentence to the paragraph level has proven a difficult matter. There are no grammars of paragraph structure. There is no general linguistic theory of the parts of speech of discourse. How do you build a coherent discourse? What basic elements govern the structure? What are the elements of the problem?

We believe that to understand discourse you have to understand discourse structure, and that a central stumbling block is the underspecificity of multisentence language. For example, on being told that

> Zurab and Maria had a fight last night.
> Maria died this morning.

you are fully within your rights to assume that the fight somehow caused Maria's death, and that Zurab was the perpetrator. But this assumption is not mandatory. Said separately, phrased differently, or embedded in an appropriate context (say, just after the sentence "Maria was diagnosed with cancer a year ago"), these two sentences do not always give rise to the assumptions. It is their juxtaposition — the way the text is structured — that makes the implications so natural and immediate that they cannot be ignored.

It has long been noted that several discourse phenomena (such as inference arising from juxtapositions, pronoun and other reference use, quantifier scoping, shifts in focus of attention, etc.) all reflect an underlying organization of the discourse which can be described as a partitioning of the material into interrelated segments. (A more precise definition is given below.) We believe that the production and interpretation of multisentence discourse succeeds only when the discourse is properly structured and all interlocutors know the structure. If not, numerous things will go wrong: pronouns will not be resolvable to their referents, the temporal structure underlying the text will be missed, the interrelationships of the various portions of the discourse will be unclear, resulting in wrong inferences, and so on.

Any person or system producing multisentence discourse must therefore confront the problem of discourse structure, which can be posed as a set of questions:

- Since the discourse under discussion is goal-based: How do the author's communicative intentions give rise to the discourse?
- Since communication succeeds only if the reader participates: How can the author guide the reader's inferences? Or: how can the author take precautions against undesired inferences?

- Since we are interested in computer-based generation: By what process can a computer plan an effective communication?

The key insight for solving these questions is the notion of text coherence. Following the definition of [Mann & Thompson 88], we define coherence as follows:

> A discourse is coherent if the hearer knows the communicative role of each portion of it; that is, if the hearer knows how each clause relates to each other clause.

In other words, a discourse is coherent and will succeed only if it is properly structured: if (i) segments properly reflect communicative intentions, and (ii) interrelationships among segments are properly expressed (so that the reader can recognize them, run the appropriate inferences, and build up the desired structures).

We have now introduced all the key notions upon which this work rests: discourse segments, intersegment relationships, communicative intentions, and reader inferences.

Theoretical Antecedents: Work on Discourse Structure

The question of what makes discourse coherent has been studied from several perspectives, including, within Computational Linguistics and Natural Language Processing work on monologic discourse[1], two major approaches: the structuralist and the functionalist perspectives. As it turns out, the theories being developed in these two perspectives are largely complementary, and in fact they seem to be converging, hopefully toward a unified model of general (single- and multi-person) discourse.

Following typical *structuralist* analyses, such as [Kamp 81], the argument goes as follows: discourse exhibits internal structure, where structural segments are defined by semantically related material. The theories tend to concentrate on the development of formalisms for representing discourse segments. These theories are strong on the formal properties of discourse segments and on the nature of the discourse structure itself (that is, the "scaffolding" supporting the text), which usually is a tree of some form, and tend to be weakest on the actual contents of the structure, such as the precise interrelationships between segments. Some of the more influential structuralist work is Discourse Representation Theory (DRT) [Kamp 81], and that of [Polanyi 88, Reichman 85, Cohen 83].

[1]With regard to dialogue, research has focused on cooperative plan-based endeavors such as tutoring and interactive explanation. As a result, many ideas are shared with work on plan recognition [Kautz 87, Hobbs et al. 88, Charniak & Shimony 90]. Several research efforts are investigating the nature and role of participants' beliefs and intentions [Pollack 86, Cohen & Levesque 90, Grosz & Sidner 90, Lochbaum 91], and much effort is focused on the types of plans that underlie this type of discourse (see [Litman 85, Lambert & Carberry 91, Ramshaw 91]). Most of these theories postulate several levels of plans, each level handling a distinct phenomenon (discourse management, domain knowledge, etc.); the levels and their particularities are hot topics in the dialogue arena.

Extending beyond dialogue-length texts, [Van Dijk 72] discusses large-scale text organization and defines the notion of macro-structures and [Rumelhart 72] develops the idea of story grammars.

The *functionalist* argument goes as follows: discourse exhibits internal structure, where the segments are defined by communicative purpose. The theories tend to concentrate on the goals of the author and on the ways these goals are reflected in the discourse structure, often as interrelationships between segments. Often, such interrelationships are viewed as reflecting plans of one sort or another which serve the interlocutors' communicative goals. The theories are strong on the particular intersegment relations and their use as operators in planning algorithms; they tend to be weakest on the precise form of the discourse structure. This approach has a fairly long history as well; researchers going back to Aristotle [Aristotle 54] have recognized that in coherent text successive pieces of text are related in a relatively small set of particular ways. Hobbs [Hobbs 78, Hobbs 79] produced a set of relations organized into four categories, which he postulated as the four types of phenomena that occur during communication. Other categorizations of typical intersentential relations were developed by [Grimes 75, Shepherd 26, Dahlgren 88, Mann & Thompson 88], to name a few.

A combination of the structuralist and functionalist ideas is embodied in the theory of discourse developed by [Grosz & Sidner 86]. This theory describe a three-way parallel analysis of discourse into the (structuralist) organization of the utterances, the (functionalist) structure of interlocutor intentions, and the attentional state (an additional record of the referentially available objects).

Computational Antecedents: Generating Coherent Text

The evolution of structuralist and functionalist approaches to discourse structure is fairly recent. Early computational work on multisentence text simply ignored the issue of text structure per se. Generators followed "guided consumption" strategies for deciding what material to include and how to organize it, such as hill-climbing (KDS) [Mann & Moore 81] or proceeding according to the structure of the domain semantics (e.g., TALESPIN [Meehan 76] and PROTEUS [Davey 79]). Early parsers either used predefined large-scale knowledge structures that spanned the relevant content of the text, such as scripts (SAM [Cullingford 78], FRUMP [DeJong 79], BORIS [Dyer 83]), or they dynamically built up structures using rules particular to the purpose, such as the argument structure work of [Birnbaum et al. 80] and [Sycara 87].

One of the first text generators that took discourse structure into account in any way was TEXT [McKeown 85]. Each schema in its library was a predefined representation of a stereotypical paragraph structure which acted as a template to mandate the content and order of the clauses in the paragraph; coherence was achieved by the correct nesting and filling-in of schemas. TEXT used four schemas – Identify, Describe, Compare&Contrast, and Attributive — to generate short texts describing various naval objects such as submarines. An example schema is shown in Figure 1. Each of its parts is defined in terms of a rhetorical predicate, which specifies what type of material may

Identification Schema

```
Identification (class & attribute/function)
{Analogy/Constituency/Attributive/Renaming}*
Particular-illustration/Evidence+
{Amplification/Analogy/Attributive}
{Particular illustration/Evidence}
```

Example

"Eltville (Germany) 1) An important wine village of the Rheingau region. 2) The vineyards make wines that are emphatically of the Rheingau style, 3) with a considerable weight for a white wine. 4) Taubenberg, Sonnenberg and Langenstuck are among vineyards of note." [PATERSON 80]

Figure 1: The IDENTIFICATION schema from TEXT, [McKeown 85].

fill that part by providing semantic attributes the material (represented in a knowledge base) must contain. The variation afforded by McKeown's schemas was extended by [Paris 87], who developed methods of tagging parts of schemas for appropriateness to various levels of sophistication of the hearer.

Though schemas remain a very clear and popular method of generating multisentence texts today (see for example [Rambow & Korelsky 92]), their utility is limited because of their essential shortcoming: the lack of representation of the purpose of each part in the whole. Without such information, the system cannot replan any portion of its text in the case that a portion should not communicate successfully, and cannot motivate why it said what it said. This shortcoming is crippling to any system that must be able to assemble its text dynamically and then reason about it, such as interactive explanation generators or documentation generators (see [Moore 89]).

In order to address this shortcoming, a method of dynamically assembling coherent discourses from basic building blocks had to be developed.

3.2 The Initial Text Planning Experiment

The planning of multisentence paragraphs by computer requires both a sound theory of text organization and an algorithm that can make efficient use of it. One of the most influential theories of text structure is Rhetorical Structure Theory (RST) [Mann & Thompson 88, Mann & Thompson 86],

which postulates that a set of approximately 25 relations suffices to represent the relations that hold within normal English texts. The study involved some hundreds of paragraphs (ranging over advertisements, scientific articles, letters, newspaper texts, etc.). The theory holds that the relations are used recursively, relating ever smaller segments of adjacent text, down to the single clause level; it assumes that a paragraph is only coherent if all its parts can eventually be made to fit into one overarching relation. Most relations have a characteristic English cue word or phrase which informs the hearer how to relate the adjacent clauses; larger blocks of clauses are then related similarly, so that eventually the role played by each clause can be determined with respect to the whole.

In order to address some of the shortcomings of schemas, the author and colleagues have carried out an investigation into the planning and generation of multisentential paragraphs over the last four years. In the experiment, some relations from Rhetorical Structure Theory were operationalized as plans and used in a text structure planner (a simplified top-down incremental refinement system patterned on NOAH [Sacerdoti 77]). The structurer planned coherent paragraphs in several domains to achieve communicative goals for affecting the hearer's knowledge in some way. The system operated between some application program (such as an expert system) and before the sentence generator Penman (see [Penman 89, Mann & Matthiessen 83]). From the application system, the structure planner accepted one or more communicative goals along with a set of clause-sized input entities that represented the material to be generated. It assembled the input entities into a tree that embodied the paragraph structure, in which nonterminals were RST relations and terminal nodes contained the input material. It then traversed the tree, submitting the input entities to Penman. A short review of the structuring process occupies the rest of this section; it is described in much more detail in [Hovy 88, Hovy 90a].

The structurer's plans were formulations of RST relations. Each relation/plan has two primary parts, a *nucleus* and a *satellite*, and recursively relates some unit(s) of the input, or another relation (cast as nucleus), to other unit(s) of the input or another relation (cast as satellite). (A simple relation/plan, SEQUENCE, is shown in Figure 2[2]). In order to admit only properly formed relations,

[2]The contents of this relation/plan can be paraphrased as follows: The plan, when used successfully, guarantees that both speaker and hearer will mutually believe that the relationship SEQUENCE-OF holds between two input entities (that is to say, that one entity follows another in temporal, ordinal, or spatial sequence). That is the contents of the RESULTS field. In order to ensure proper ordering and focus, one input entity is bound to the variable ?PART in the sc Nucleus requirements field and the other to the variable ?NEXT in the SATELLITE REQUIREMENTS field. No other semantic requirements hold on the input entities individually. There is, however, the requirement that they be semantically related by some kind of sequential link (in the current domain, the temporal relation NEXT-ACTION); that is, that ?PART does in fact precede ?NEXT; this requirement is stated in the NUCLEUS+SATELLITE REQUIREMENTS field. Suggestions for additional input material related to the nucleus are contained in the NUCLEUS GROWTH POINTS field: these call for circumstantially related material (time, location, etc.), attributes (size, color, etc.) and purpose. They are stated in terms of mutual beliefs in order to act as subgoals that the planner must try to achieve. A similar set is associated with the satellite. The typical order of expression in the text is nucleus first and the satellite, using

nuclei and satellites contained requirements that had to be matched by characteristics of the input. In addition, nuclei and satellites contained *growth points*: collections of goals that suggested the inclusion of additional material in the places where it was typically placed (such as discussed, for example, in [Conklin & McDonald 82]). Determining the contents of growth points was a major task; in one domain, for example, not only were dozens of paragraphs analyzed, but the expert responsible for producing them was interviewed and taped over a period of three days.

On finding (an) RST relation/plan(s) whose effects include achieving (one of) the system's communicative goal(s), the structure planner searched for input entities that matched the requirements holding for its nucleus and satellite. If fulfilled, the planner then considered the growth points of the relation/plan. It tried to achieve each newly activated growth point goal by again searching for appropriate relation/plans and matching their nucleus and satellite requirements to the input, recursively, adding successfully instantiated relations to the paragraph tree structure. The planning process bottomed out when either all of the input entities had been incorporated into the tree or no extant goals could be satisfied by the remaining input entities. The tree was then traversed in a depth-first left-to-right manner, adding the relations' characteristic cue words or phrases to the appropriate input entities, and transmitting them to Penman to be generated as English clauses.

The paragraph structure planner was been applied to several domains, including expert systems (see [Hovy 88]), a code development system (see [Hovy & Arens 91], and a multimodal database information display system [Arens et al. 88]. We take here an example from the latter, the Integrated Interfaces program, a multimodal system that uses maps, tables, and paragraphs of text to answer users' requests for the display of information from a data base of naval information about ships' deployments. In the example, the display planner furnishes the RST text structure planner with a set of six related entities, along with the following goal:

```
(BMB SPEAKER HEARER (SEQUENCE-OF E1 ?NEXT)
```

This goal can be paraphrased as: achieve the state in which both the speaker and the hearer mutually believe that input entity E1 is followed in time by some other input entity[3]. After rewriting the input into a standard form (called here input entities, and shown in Figure 3), the structurer proceeds to plan a paragraph, producing the tree shown diagrammatically. It then traverses the tree and supplies the input entities at the leaves to Penman to be generated as sentences, one by one. For continuity of exposition, similar navy examples will be used throughout this document.

The problem of planning coherent text can be characterized in specific terms as follows. Assuming that input elements are sentence- or clause-sized chunks of representation, the permutation

either no cue word or one of "then" or "next".

[3]The term BMB stands for *believe mutual belief*, and is taken from [Cohen & Levesque 85], who develop a notation for reasoning about beliefs and mutual beliefs during the communication of speech acts. This terminology is discussed in more detail in Section 3.3.2.

Figure 2: The RST relation/plan SEQUENCE

```
Name: SEQUENCE

Results:
  ((BMB SPEAKER HEARER (SEQUENCE-OF ?PART ?NEXT)))

Nucleus requirements/subgoals:
  ((BMB SPEAKER HEARER (TOPIC ?PART)))

Satellite requirements/subgoals:
  ((BMB SPEAKER HEARER (TOPIC ?NEXT)))

Nucleus+Satellite requirements/subgoals:
  ((NEXT-ACTION ?PART ?NEXT))

Nucleus growth points:
  ((BMB SPEAKER HEARER (CIRCUMSTANCE-OF ?PART ?CIR))
   (BMB SPEAKER HEARER (ATTRIBUTE-OF ?PART ?VAL))
   (BMB SPEAKER HEARER (PURPOSE-OF ?PART ?PURP)))

Satellite growth points:
  ((BMB SPEAKER HEARER (ATTRIBUTE-OF ?NEXT ?VAL))
   (BMB SPEAKER HEARER (DETAILS-OF ?NEXT ?DETS))
   (BMB SPEAKER HEARER (SEQUENCE-OF ?NEXT ?FOLL)))

Order: (NUCLEUS SATELLITE)
Relation-phrases: ("" "then" "next")
Activation-question:
  "Could ~A be presented as start-point, mid-point,
   or end-point of some succession of items along
   some dimension? -- that is, should the hearer
   know that ~A is part of a sequence?"
```

```
((ENROUTE E1)
 (ACTOR E1 K1)
 (DESTINATION E1 S1)
 (NEXT-ACTION E1 A1)
 (LOCATION E1 P1))
((ARRIVE A1)
 (ACTOR A1 K1)
 (TIME A1 T1))
 (NEXT-ACTION A1 L1))
((LOAD L1)
 (ACTOR L1 K1)
 (STARTTIME L1 T2)
 (ENDTIME L1 T3))
((SHIP K1)
 (NAME K1 KNOX)
 (READINESS K1 C1)
((PORT S1)
 (NAME S1 SASEBO))
((READINESS-STATUS C1)
 (NAME C1 C4))
((POSITION P1)
 (HEADING P1 H1)
 (LATITUDE P1 79)
 (LONGITUDE P1 18))
((HEADING H1)
 (COURSE H1 195))
((DATE T1)
 (DAY T1 24)
 (MONTH T1 4))
((DATE T2)
 (DAY T2 25)
 (MONTH T2 4))
((DATE T3)
 (DAY T3 28)
 (MONTH T3 4))
```

```
                 |
                 |
             SEQUENCE
              /    \
             /      \
         CIRC        SEQUENCE
         /  \          /  \
     ATTR    ATTR     A1   L1
     /  \    /  \
   E1   C1  P1   H1
```

Knox, which is C4, is en route to Sasebo. It is at 79N 18E heading SSW. It will arrive on 4/24, and will load for four days.

Figure 3: Example of navy data base assertions input to the structurer, the resulting paragraph structure tree, and corresponding text (left branches of the tree are nuclei, right branches, satellites).

set of the input elements defines the space of possible paragraphs. A simplistic, brute-force way to achieve coherent text would be to search this space and pick out the coherent paragraphs. This search would be factorially expensive; for example, for the navy paragraph above, the 6 input entities provide 6! = 720 possible paragraphs. By by utilizing the constraints imposed by coherence, one can formulate operators out of the coherence relations that guide the search and significantly limit the search to a manageable size. In the example, the relation/plan operators produced fewer than five candidate paragraphs; the best one was selected using a simple evaluation metric based on the number of unused input entities and the balance of the tree.

This experiment was an early step toward the eventual ability to plan coherent discourse dynamically. Capturing the internal organization and rhetorical dependencies between clauses in the text, the paragraph structure tree enables some powerful reasoning about the text. For example, since it contains the derivation of each part of the paragraph, one knows the role each clause plays with respect to the whole, and thus can identify and repair mistakes. In addition, when the text structure is known, various other sentential aspects can be determined. Note in the example text the following:

- realization of the satellites of the ATTRIBUTE relation as relative clauses: `Knox, which is C4...`instead of, say, `Knox is C4. It is en route....` This realization pattern for the ATTRIBUTE satellite is standard in English.
- use of the future tense in the final sentence. Since information provided by the data base was always based on the present time, anything that appeared in the satellite of a temporal SEQUENCE relation had to be in the future.
- linking of the last two clauses into a single sentence. Deciding to link clauses is easily done when a paragraph structure is available; the complexity of each subtree can readily be determined and appropriate sentence-building decisions made.

3.3 Resulting Issues and Research Experiments

The initial experiment established that is is possible to plan coherent paragraphs for a variety of domains using RST relations as plan operators. But it also opened up a set of unresolved problems that must be addressed before discourse planning can become a reality. While for example it is clear that such relations play a central role in understanding and generating discourse, their precise nature and uses had to be uncovered. The major aspects of the problem can be broken down into following seven issues:

1. Discourse structure
2. The content and format of plans

3. A collection of relations/plans

4. Predefined structures (schemas)

5. Controlling planning by focus shift

6. Planning on the sentence level

7. Discourse relations and text formatting

Most of these issues have been addressed in subsequent experiments by the author and by others; none have been fully solved. Taken together, however, the current state of text planning work represents a significant advance over what was known about the automated planning and generation of discourse even five years ago.

3.3.1 Discourse Structure

As mentioned earlier, the nature of the discourse structure is still being debated. Most descriptions are based on a mixture of intuition, arguments from linguistic studies, observations from conversational analysis, as so on. Instead of adopting any of these theories and so deciding beforehand what the discourse structure should be, the approach taken in this work was pragmatic: use only what is required to produce coherent and fluent English text.

Synthesizing the results of computational experiments in a variety of domains by several researchers (aside from the author, [Moore & Swartout 90, Paris 90, Maybury 90, Cawsey 90, Dale 88] and others) and taking into account the theoretical work, the following general assertions about the structure of plan-based English discourse have crystallized out:

1. **Discourse:** A discourse (a text) is a structured collected of clauses. By their semantic relatedness, clauses are grouped into segments; the discourse structure is expressed by the nesting of segments within each other according to specific relationships. A discourse can thus be represented as a tree structure, in which each node of the tree governs the segment (subtree) beneath it. At the top level, the discourse is governed by a single root node; at the leaves, the basic segments are single grammatical clauses.

2. **Purpose:** Each discourse segment has purpose, which (following [Grosz & Sidner 86]) we call the Discourse Segment Purpose (DSP) and represent at each node of the tree. Each DSP is a communicative goal of the speaker. In a successful discourse, the contents of each segment achieve its DSP. Each segment can thus be seen as a step in a plan to achieve the overall communicative purpose of the discourse.

3. **Coherence:** In any plan, the steps are ordered or partially ordered due to underlying interrelationships and dependencies among their contents. These interrelationships must be respected in order to achieve the plan successfully. In language, a discourse in which the reader knows how each portion relates to its neighbors and thus to the whole is called *coherent.* Coherence is a hallmark of a successful discourse and is enforced by discourse structure relations such as RST relations (see Section 3.3.2).

4. **Definition:** A discourse segment S is a tuple (*name, purpose, content*), where:
   - The *name* is a unique identifier for the segment.
   - The *purpose* is one or more communicative goals the speaker has with respect to the hearer's state of knowledge, opinion, goals, etc.
   - The *content* is either:
     - an ordered list of discourse segments, together with one or more discourse segment relations that hold between them (either there is a relation between every two adjacent segments in the list, or a relation holds among all the segments in the list simultaneously); or
     - a single discourse segment; or
     - the semantic material to be communicated (usually using a single clause). This material usually takes the form of a set of knowledge base assertions or data base facts.

5. **Definition:** A discourse structure D is a discourse segment which is not contained in any discourse segment and all of whose leaves (the innermost segments) contain semantic material to be communicated.

This formulation of discourse segment and discourse structure is purposely rather general, in order to accord with that of [Grosz & Sidner 86, Asher 91], and [Polanyi 88], as well as with the work on intention recognition [Allen & Perrault 80, Litman 85, Pollack 86, Lambert & Carberry 91]. The RST based paragraph trees from the first experiment (Section 3.2) can be reformulated to conform to this definition, by the addition of explicit communicative goals to each relation (for presentational clarity, however, this has not been done in this paper). Similarly, with minor reformulation, the text structures built by the planners EES [Moore 89], EPICURE [Dale 88], TEXPLAN [Maybury 90], SPOKESMAN [Meteer 90], POPEL [Reithinger 91], and similar can be cast in this form as well.

3.3.2 Plan Content and Format

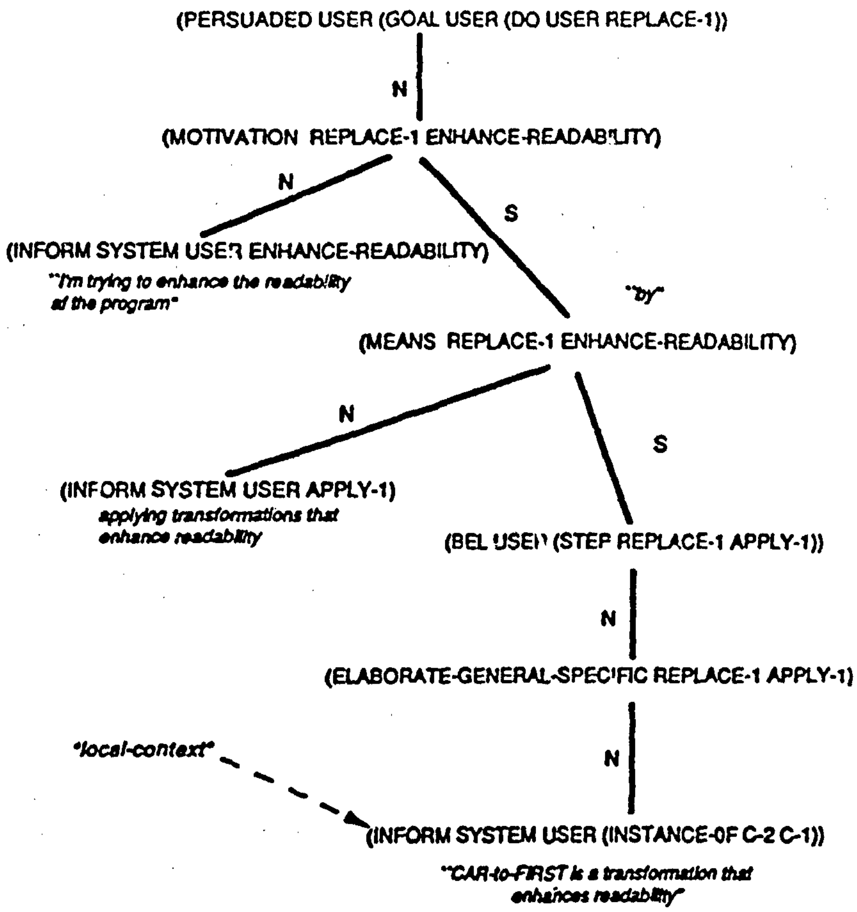
What kinds of plans are needed to generate coherent text? This question, still a long way from fully answered, has received much attention in the text planning community.

Discourse Plans vs. Intersegment Relations

Since the first experiment with RST-based text structure planning, the nature of the relation/plans has been an issue. In Rhetorical Structure Theory [Mann & Thompson 86], relations are structural entities that reflect underlying semantic and interpersonal relationships between the discourse segments. However, for use in the RST structurer, the relations had to be viewed as plans — the operators that guided the planner's search through the permutation space of inputs. The structurer's goals were all directly related to its relations, meaning that it was limited to a "rhetorical" language, planning to achieve goals such as "create an elaboration between the current material and some additional material".

As pointed out in [Moore & Swartout 90], employing relational terms as goals seems misplaced; the structurer conflates intentional with "rhetorical" (i.e., structural) information. Moore, Paris, and Swartout set out to develop a new plan language and a new set of plan operators, eventually incorporating them in a text planner they built for the Explainable Expert System (EES) [Moore & Swartout 90, Moore 89, Moore & Paris 89]. The EES planner contained such plans as INFORM, RECOMMEND, INFORM-AND-PERSUADE, PERSUADE-BY-MOTIVATION, MOTIVATE-ACT-BY-MEANS as general domain-independent operators and such plans as PERSUADE-INSTANCE-IMMEDIATE-SUBCLASS-OF-REDESCRIPTION, BY-MEANS-COMPLEX-METHOD, BY-MEANS-SIMPLE-METHOD as somewhat more domain-specific operators. In addition, the EES planner contained several RST-like plan operators, including SEQUENCE-STEPS, CONTRAST, ELABORATE-OBJECT-ATTRIBUTE, ELABORATE-GENERAL-SPECIFIC, ELABORATE-PROCESS-STEP. Using this plan language, Moore, Paris, and Swartout planned discourse structures that contained terms of a more "intentional" nature. An example of some of the EES plans and of a discourse structure fragment appear in Figure 4. A similar approach was followed by Maybury in his TEXPLAN planner [Maybury 90]. He, too, used a mixture of domain-independent, domain-dependent, and RST-like operators.

Neither approach is wholly satisfactory. Certainly, for plan-based discourse, the plans employed should express the author's communicative intentions. But by adding RST-like plan operators into their plan libraries, Moore and Paris and Maybury undercut their own argument, since their planners also then plan with relation/plans at various points, usually toward the leaves of the discourse structures. The dilemma is resolved when one recognizes that the two types of object — intentional plans and discourse relations — perform different functions and hence are needed *simultaneously* to govern the discourse. In order to determine what material to include and to provide the overall structure of the discourse, intentional plans are most appropriate; within this framework, it is the function of structural relations to ensure textual coherence, prevent unintended inferences, govern sentence formation, tense, pronominalization, focus shift, etc. (see subsequent sections). To see this, note that the same communicative purpose can be achieved in many ways; for example, the goal to PROVE clause (1) can be achieved using several discourse relations with




```
EFFECT:        (EVIDENCE (RESULT ?G ?value))
CONSTRAINTS:   (AND (GOAL S ?G)
                    (METHOD-SELECTED ?method ?G)
                    (RESULT ?G ?value))
NUCLEUS:       (INFORM S H (METHOD-SELECTED ?method ?G))
SATELLITE:     ((ELABORATE-GENERAL-SPECIFIC
                    (METHOD-SELECTED ?method ?G))
                    *required*)
```

Figure 4: Example discourse structure and text plans from the EES planner, from [Moore 89].

clause (2):

> CAUSE: "(1) *He knows how to deal with red tape because* (2) *he lives in Moscow.*"
> CIRCUMSTANCE-LOCATION: "*Living in Moscow, he knows how to deal with red tape.*"
> SEQUENCE-TIME: "*After he went to live in Moscow, he knew how to deal with red tape.*"

In general, some text genres tend to be more intentional (discourse analyses of explanatory discourse, etc.) and others less so and more structural (such as encyclopedia entries). In the former, intentional plans dominate, while in the latter, large subportions of the discourse serve a single discourse intention (usually, DESCRIBE) and are governed by a considerable tree of discourse relations (texts generated by TEXT [McKeown 85] and the RST structurer are of this type; the main intentional goal is to describe). The definition of discourse segments in Section 3.3.1 prescribes both intentions and structural relations for this reason.

Differentiating the two types of object into intentional plans and structural relations seems to correspond with the distinction made in [Austin 65] between sentences with perlocutionary effect (such as persuading, motivating, etc.) and those with illocutionary effect (such as elaborating, identifying, describing, etc.), though, as Maybury's attempt to do so shows, this distinction is unfortunately hampered by the vagueness of the notions of perlocution and illocution and the imprecision of plans' and relations' definitions.

More detailed arguments for the nature and need of intentional plans appear in [Moore & Paris 91, Moore 89, Paris 90].

A Formalism for Relation/Plans

In the initial experiment, RST relations were operationalized as plans in a straightforward manner. The formalization was found to be inadequate for explanatory discourse, however, prompting Moore, Paris, and Swartout to define for the EES text planner plans that include, in addition to the operator effect, nucleus, and satellite fields also a field for *constraints* — the facts (within the system's knowledge base or user model) that had to be true about the data before the plan could be applied. Maybury further elaborated the formalism, adding also *preconditions* of two kinds, *essential* and *desirable* (an example of this formalism for text plans is shown in Figure 5; note that the entries for the DECOMPOSITION field are ordered and, unless explicitly flagged, mandatory subgoals, and that planning proceeds along the HEADER fields, not the EFFECTS — that is, subgoals are achieved by plans whose HEADER fields match; the EFFECTS are simply for updating the hearer model).

Based on the above work, as well as on the EDGE planner [Cawsey 90] and the plan representation in SPOKESMAN [Meteer 90], we define a plan P as a tuple (*name effects constraints preconditions decomposition*), where:

```
NAME              extended-description
HEADER            Describe(S, H, entity)

CONSTRAINTS       Entity?(entity)
PRECONDITIONS
  ESSENTIAL       KNOW-ABOUT(S, entity) ∧
                  WANT(S, KNOW-ABOUT(H, entity))
  DESIRABLE       ¬ KNOW-ABOUT(H, entity)

EFFECTS           KNOW-ABOUT(H, entity)

DECOMPOSITION     Define(S, H, Entity)
                  optional(Detail(S, H, entity))
                  optional(Divide(S, H, entity))
                  optional(Illustrate(S, H, entity)) ∨
                           Give-Analogy(S, H, entity))
```

Figure 5: Text plan EXTENDED-DESCRIPTION from [Maybury 90].

- The *name* is a unique identifier of the segment.
- The *effects* are one or more communicative goals that the plan achieves, if properly executed. These goals pertain to the speaker's desire with respect to the hearer's state of knowledge, opinion, goals, etc.
- The *constraints* are facts in the knowledge base or the user model that must hold before the plan may be used.
- The *preconditions* are facts in the knowledge base or user model that should hold for felicitous communication. If they are violated, the hearer may be confused, and (in a dialogue situation) the planner should mark preconditions it violates, in order to facilitate locating what to repair when things go wrong.
- The *decomposition* is an unordered list of subgoals to be achieved. Each subgoal may be flagged as optional, in which case the planner can ignore it under appropriate conditions (conditions depend on the sophistication of the planner: at the minimum, it can simply ignore the subgoal if instructed to produce terse text; being more sophisticated, the planner may reason about various contributing factors, such as the balance of material within the discourse structure so far, the level of detail of the indicated material, etc.). Ordering is achieved by structuring with discourse relations.

Since the communicative intentions of the author are (usually) related to the reader, these in-

tentions, the plans, their preconditions, etc., must be formulated in terms of beliefs, knowledge, opinions, etc. Suitable terms for this purpose are provided by the formal theory of rational interaction being developed by, among others, Cohen, Levesque, and Perrault. For example, in [Cohen & Levesque 85], Cohen and Levesque present a proof that the indirect speech act of requesting can be derived from the following basic modal operators:

- **(BEL x p)** — **p** follows from **x**'s beliefs
- **(BMB x y p)** — **p** follows from **x**'s beliefs about what **x** and **y** mutually believe
- **(GOAL x p)** — **p** follows from **x**'s goals
- **(AFTER a p)** — **p** is true in all courses of events after action **a**

as well as from a few other operators such as **AND** and **OR**. They then define *summaries* as, essentially, speech act operators with activating conditions (*gates*) and *effects*. These summaries closely resemble, in structure, the plans developed in text planners, with gates corresponding to constraints on material and effects to intended effects. Most text planners at this time use modal operators of belief along these lines.

3.3.3 A Library of Relation/Plans

The Problem: Which Relations? How Many?

One of the central problems confronting discourse and text planning work is the nature of the intersegment relations: are they semantic, "rhetorical", intentional, or what?

Approaching the problem of discourse structure from several intellectual subfields, various researchers have produced lists of intersegment relations — from philosophers (e.g., [Toulmin 58]) to linguists (e.g., [Quirk & Greenbaum 73, Halliday 85]) to computational linguists (e.g., [Hobbs 79, Mann & Thompson 88]) to Artificial Intelligence researchers (e.g., [Schank & Abelson 77, Moore 89, Dahlgren 88]). Typically, their lists contain between five and thirty relations, and they argue that (at least) tens of interclausal relations are required to describe the structure of English discourse; we call this the *Profligate Position*.

On the other hand, some researchers, (e.g., [Grosz & Sidner 86, Polanyi 88, Kamp 81]) prefer not to identify a specific set of such relations. They argue that trying to identify the "correct" set is a doomed enterprise, because there *is* no closed set; the closer you examine intersegment relationships, the more variability you encounter, until you find yourself on the slippery slope toward the full complexity of semantics proper. Though they do not disagree with using relationships between adjacent text segments to provide meaning and enforce coherence, they object to the notion that some small set of relations describe English discourse adequately. As a counterproposal, Grosz and

Sidner define two basic relations, DOMINANCE and SATISFACTION-PRECEDENCE, which carry intentional (that is, goal-oriented, plan-based) but no semantic import, and suffice to represent tree-like nature of discourse structure. We call this the *Parsimonious Position.*

Collecting Relations

While the parsimonious relations may satisfactorily represent discourse structure for purposes of analysis, practical text generation experience, such as [McKeown 85, Hovy 88, Moore & Swartout 90, Paris 90, Rankin 89, Cawsey 90, Maybury 90, Dobeš & Novak 91], has shown that they are insufficient and that planners need considerably more information of rhetorical and semantic nature in order to ensure successful communication. For example, when generating the following two clauses

"His car was much admired because it was a red Ferrari."

the author needs to know more than the relationship of the intentions underlying each clause. He or she also needs to know which semantic interrelationship to express: it is the semantic relation of causality that provides the appropriate linking word and much of the structural/realizational information (had the interclausal relationship been temporal coincidence, the cue word would have been "when"; had it been elaboration, the second clause would have been subordinated to the first in a relative clause "His car, which was...", and so on).

Accordingly, in 1989 the author started collecting intersegment relations that are expressive enough to satisfy the requirements of text planning systems while avoiding an unbounded ad hoc collection of semantic relations. Over 350 such relations from approximately 30 researchers in various fields were collected and taxonomized; see [Hovy 90b]. Subsequently, in joint work, over 50 additional relations in other sources were found and an improved taxonomization, consisting of about 70 relations, was produced; see [Maier & Hovy 91]. The relations, organized into a taxonomy, are reproduced in Figure 6 and described in more detail in [Hovy & Maier 92].

Of course, there is no guarantee that the relations collected are indeed the "right" and only ones. Their strongest support is that they are the amalgamation and synthesis of the efforts and proposed terms of several investigations in different fields, including actual attempts to construct working text planners and discourse analyzers. When different interclausal relations are proposed, we expect that the hierarchy will grow primarily at the bottom, and that the ratio of the number of relations added at one level to the number of relations added at the next lower level will be low, for all levels. This accords with our experience when compiling the hierarchy: halfway through this study, the topmost tiers had essentially been established, and almost all new relations found were simply specializations of existing ones.

Taxonomizing the Relations

Given the semantic overlaps of many of the relations, a natural taxonomy soon suggested itself: a two-dimensional hierarchic organization by increasing semantic specificity, with one dimension

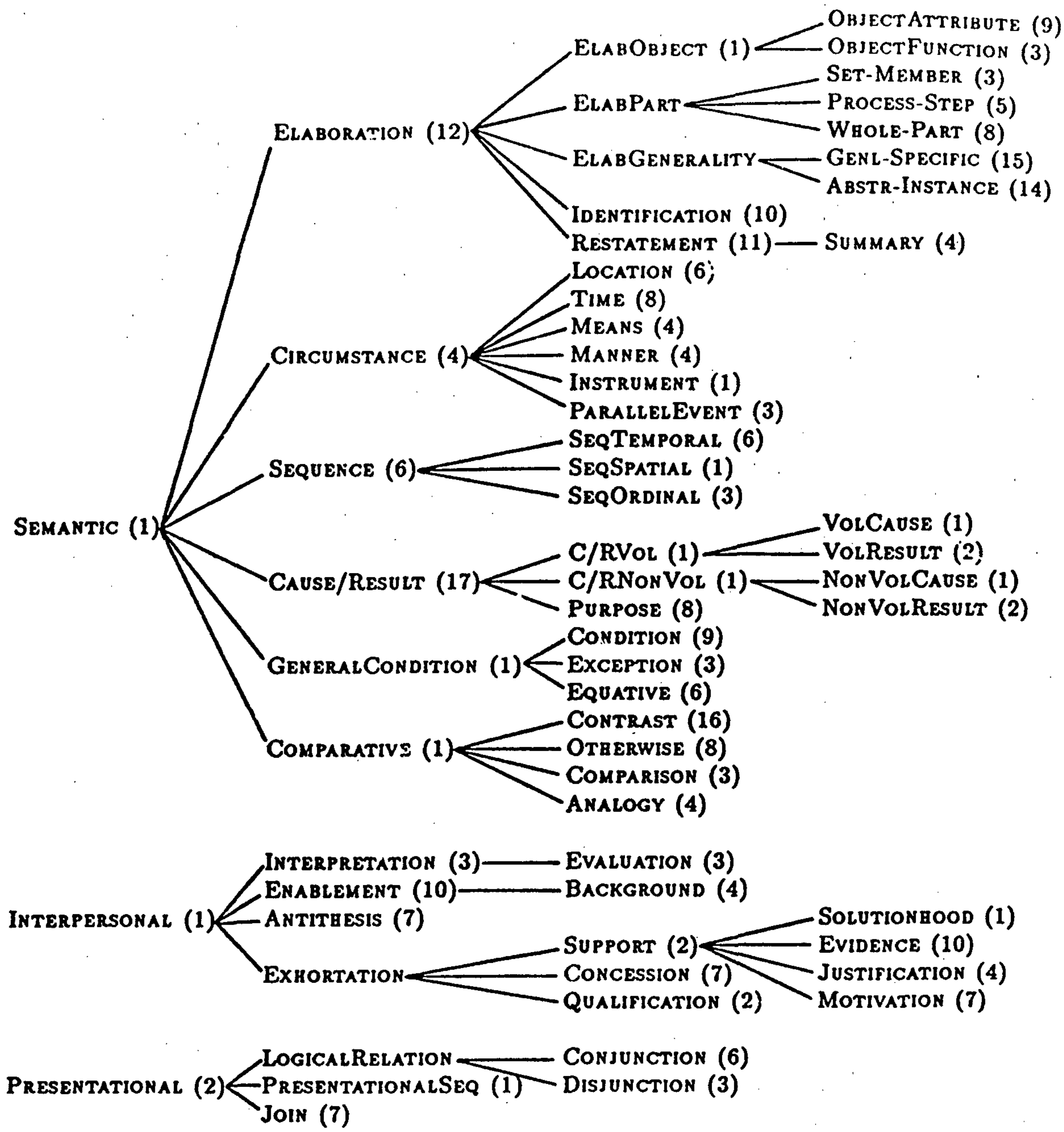


Figure 6: A Hierarchy of Intersegment Relations. The number associated with each relation indicates the number of different researchers who have listed the relation and may be interpreted as a vote of confidence in it.

constrained in the number of relations and the other unconstrained (the more a relation is specified to distinguish it from others, the more its semantics are enhanced — adding semantic features is the nature of increasing specification — and the lower it appears in the hierarchy). Though the unboundedness at the bottom places one on the slippery slope toward having to deal with the full complexity of semantic meaning, there is no reason to fear such complexity The terms are well-behaved and subject to a pattern of organization which makes them manageable: all the pertinent information about discoursal behavior is captured near the top; each relation inherits from its ancestors all necessary processing information, such as cue words and realization constraints, and adds its unique peculiarities, to be used for inference (in parsing) or for planning out a discourse (in generation). Increasing differentiation of relations, continued until the very finest nuances of meaning are separately represented, need be pursued only to the extent required for any given application.

Our top-level classification into three (see Figure 6) is motivated by several factors. First, our view of generation as essentially a planning process fosters a functional perspective on language and on the relations in particular, We therefore partitioned the relations into three broad groups according to which primary function they perform in text. (A similar subcategorization strategy was discussed in [Mann & Thompson 88]). The three functions themselves are motivated by Halliday's subcategorization of linguistic phenomena into three so-called metafunctions *ideational* (i.e., semantic), *interpersonal* (i.e., author- and/or addressee-related), and *textual* (i.e., presentational) [Halliday 85]. A second reason is the difference in relations' illocutionary force. All the ideational relations are expressed by the single illocutionary act DESCRIBE, while the interpersonal relations are expressed by various perlocutionary acts, including CONVINCE, MOTIVATE, and JUSTIFY. (See [Maybury 90, Maier & Hovy 91] for discussions.)

In conjunction with this taxonomizing work, we are currently collecting attempts to provide precise, formal definitions of these relations, for example from [Mann & Thompson 88, Ivir et al. 80, Hobbs 79, Hobbs 90, Sanders et al. 91, Martin 92, Lascarides & Asher 91, Sanders et al. 91].

3.3.4 Schemas

It has become clear, from several attempts at planning longer texts, that systems without some explicit control over the development of larger spans of text than a single paragraph are not feasible in practise. There is simply too much variability in text plans or discourse structure relations that must support flexible text structure planning. Rather, as argued in for example [McKeown 85, Mann 87, Rambow 90, Mooney et al. 90], one should capture the idiosyncratic regularities of discourse structure, which may depend on genre, domain, or even just custom, in schemas and use them as frozen plans by simple instantiation. In those places where additional structuring is required — when no frozen plan exists to achieve the communicative intention - then

discourse structure plans and relations should be used.

Fortunately, it is possible to formulate schemas as fossilized discourse structures and discourse structure relation/plans as mini-schemas, providing a homogeneity of representation that simplifies the planning process. A way of melding the two techniques was outlined in [Hovy 90a], by exercising appropriate control over optional additional material (the material, to use the above terminology, whose inclusion and order is captured in the growth point goals). By treating growth point goals as *injunctions* that specify the type and order of additional material to include, rather than as *suggestions* to do so, a relation/plan is a schema instead of a plan proper. Of course, some growth point goals can be made required and others optional, enabling relation/plans simultaneously to incorporate both fixed structural options that are not justified by reasoning (i.e., act as schemas), as well as relational patterns that are developed dynamically (i.e., support opportunistic planning). This hybrid approach combines the complementary strengths of schemas and plans (the former being simple and easy to use and the latter supporting dynamic extensibility).

This treatment has been adopted in some form or another by most newer text structure planners: both the EES and the TEXPLAN planners, for example, label subgoals to be achieved in their plans either *optional* (in which case they act as suggested growth points) or not (the default; in which case they are treated as schema entries); see Figures 4 and 5 and [Moore & Swartout 90, Maybury 90].

Several open issues remain. There is as yet no representation for schemas that captures also the underlying semantic and rhetorical interrelations of the parts. Also, when growth point goals are treated as suggestions for additional growth, two problems are immediately introduced:

- Which growth point goals should be considered?
- In what order should new growths be added to the discourse?

It is easy to think of criteria for controlling the inclusion, but difficult to formalize them adequately; for some candidates see [Hovy 90a]. One criterion, however, has been studied to some degree. This is the effect of theme development and focus shift on discourse structure, and to it we turn next.

3.3.5 Focus Shift

In any plan, the sequence of steps may be fixed or not, depending on the underlying interrelationships among their contents. In general, there is no way to tell a priori how the parts of a plan must be ordered before they have been instantiated with actual material. This means that ordering requirements usually cannot be precompiled into plans, which means that some additional mechanism has to provide additional control. This is not surprising; coherence is a not unitary phenomenon, capturable simply in a single knowledge structure; it results from the confluence of a number of considerations.

One such consideration is focus. This section describes an experiment to control the discourse planning by using focus shift constraints as decision criteria[4]. Focus we define as the location of the principal inferential effort needed when understanding the text[5]. Linguistic investigations reveal that there are strong constraints on what material may occupy the focus position as a text progresses, rules which have been computationalized and used by [Sidner 83, McKeown 85]. In our experiment, we used the technique of Focus Trees to manage allowable shifts of the focused object, as developed at the University of Delaware [McCoy & Cheng 88, McCoy 85]. The text structure planner constructed the paragraph structure and a Focus Tree in tandem. During the expansion of a node in the paragraph structure, the structurer applied all the growth point goals active at that point and collected the resulting candidate relations and their associated clause-sized input entities. Each candidate growth entity was then checked against the currently allowed focus shifts in the Focus Tree, and invalid candidates were simply removed from consideration. In general, one of three possibilities ensues:

1. Only one candidate remains. In this case, growth proceeds straightforwardly with this candidate.

2. More than one candidate remains. In this case all candidates are coherent based on rhetorical structure and focus but additional measures, still to be developed, must be employed to select the best of these. (As an interim practical solution, the growth points in the plan can be ordered by typical occurrence.)

3. No candidates remain. In this case, depending on the overall stylistic goals of the system, two options ensue:

   (a) Tree growth is simply stopped.

   (b) Tree growth continues at this point, in the default order as above, but the text is linguistically marked to indicate a focus shift. Typically, this involves reordering segments of the discourse structure to ensure adherence to focus shift constraints as well as generating appropriate surface forms.

A brief example to illustrate the point: to produce the paragraph in Figure 3, the structure planner treated growth points as injunctions, fixing the order of appearance. When this requirement was lifted, the structurer built many more paragraph structures using the same material, including the structure shown in Figure 7 (a)). This structure was made acceptable to the Focus Tree criterion by reordering the C4 clause to precede the **enroute** clause. This involved inverting the

[4]This work was performed by the author and Prof. Kathleen McCoy from the University of Delaware.

[5]See [Hovy & Lavid 92]. Severe terminological confusion surrounds the issue of focus, theme, and given; we take focus here in the sense of the Prague School [Daneš 74] and [Fries 81, Halliday 67] to mean a privileged element of the clause that usually appears in its latter, high-informational, portion.

```
                 (a)                                        (b)
                  |                                          |
                 SEQ                                        SEQ
                /   \                                      /   \
            ATTR     SEQ                            ATTR-1       SEQ
           /   \    /   \                           /   \       /   \
       CIRC     C1 A1    L1                       C1    CIRC   A1    L1
      /    \                                           /    \
    E1      ATTR                                     E1      ATTR
           /    \                                           /    \
         P1      H1                                       P1      H1
```

```
(a) Knox is en route to Sasebo.  It is at 79N 18E heading SSW.  It
    is C4.  It will arrive on 4/24, and will load for four days.

(b) With readiness C4, Knox is en route to Sasebo.  It is at 79N 18E
    heading SSW.  It will arrive on 4/24 and will load for four days.
```

Figure 7: (a) Another version of the Navy text, treating growth points in free order, and (b) using Focus Trees to ensure proper focus shifts (ATTR-1 stands for the inverse relation, in which the order of Satellite and Nucleus is switched).

ATTRIBUTIVE relation nucleus and satellite, giving a linguistically marked text by focusing on the readiness status. This work is reported in [Hovy & McCoy 89].

3.3.6 Sentence Level Planning

Even after taking into account the constraints imposed by focus, the discourse structure does not contain all the information required for the successful realization of text. One of the major open problems is the scoping of information into sentences and noun phrases. For example, the final SEQUENCE segment in Figure 3 has the following realizational alternatives:

(a). `It will arrive on 4/24 and will load for 4 days.`

(b). `It will arrive on 4/24.  It will load for 4 days.`

and on the noun phrase level the first ATTRIBUTE relation has at least:

(c). `Knox, which is C4, is en route.`

(d). `Knox is en route and it is C4.`

(e). `Knox is en route.  It is C4.`

Often, situations in which different sentence allocations exist can be recognized by characteristic configurations of the discourse structure. The ATTRIBUTE relation provides a simple example: Since it always holds between a clause constituent (such as the actor of a process) and another clause (some attribute of the actor), the satellite (the attribute) can be realized as a relative clause to the nucleus (the process containing the constituent), as long as the nucleus is not itself a subtree in the discourse. A similar problem arises with a chain of SEQUENCE relations. This problem becomes pronounced with longer chains.

Any solution on the clause level must take several issues into account: focus, the complexity of the remainder of the discourse substructure, the desired overall style of the text (such as a general preference for simple or complex sentences), the rhythm of sentences (long alternating with short, as suggested in numerous books on good style, such as [Shepherd 26]). The most concrete work on this point is a set of heuristics to govern sentence formation by Scott [Scott & De Souza 90, De Souza et al. 89]:

1. A satellite can only be embedded in its nucleus
2. Embedding can be realized as an adjective, appositive NP, PP, or relative clause, in this order of preference
3. Embedding can occur in the leftmost nuclear clause with the same focus value
4. Satellites in a LIST within an ELABORATION should be embedded, provided there are no, or else more than one, remaining clauses
5. Coordination occurs only between elements of LIST, SEQUENCE, and CONTRAST relations

6. The more shared parameters between clauses, the more they should be coordinated
7. Prefer coordinating NPs over PPs over Vs or VPs
8. Sentences should contain no more than 3 clauses
9. Sentences should contain at most one level of embedding
10. Embedding should occur before coordination and before focus transformations

Within noun phrases, the problem of delimiting and organizing content involves three major issues. The first issue relates to pronominalization. It is widely accepted that pronominalization is sensitive to segmental boundaries, at least on the relatively major level; see for example [Björklund & Virtanen 89], or the analyses of conversations by Passoneau, which suggest that discourse referents are available for pronominalization in the local context only [Passoneau 91]. Studies by [Levy 84, Marslen-Wilson et al. 82] indicate that explicit referring expressions (say, a full noun phrase instead of a pronoun) help indicate discourse segment boundaries. The availability of the discourse structure as a tree of intersegment relations, in which segments manifest themselves as subtrees, enables the development of sophisticated pronominalization strategies. Exactly which segment boundaries permit pronominalization, however, remains an open question.

The second issue arises in cases where material in a dependent clause can be realized instead within the noun phrase proper (as an adjective, say). Again from Figure 3, "Knox, which is C4,..." could have been realized as "the C4 Knox..."; in Figure 7, we deemed the clause-sized "Being C4, Knox..." (which was realized by default) unacceptable, preferring the realization "With readiness C4, Knox...". Determining the optimal syntactic class of material depends, among other things, on the balance of the paragraph structure tree, on focus, and on the stylistically desired density of information in the noun phrase.

The third issue, aggregation, appears frequently, and arises from the fact that information represented by the domain system as separate individuals is often generated as a group sharing pertinent features. For example, the Integrated Interface data base represented each ship separately, but could decide to display several ships moving together. Without rules for syntactically grouping the ships into a single clause or portion of a clause, the text was of poor quality:

```
MEKAR-87 takes place in the South China Sea from 10/20 until 11/13.
Knox, Fanning, and Whipple are participating.  Knox arrives on 10/20.
It leaves on 10/31. Fanning arrives on 10/20.  It leaves on 11/13.
Whipple arrives on 10/29.  It leaves on 11/13.
```

It is easy to invent aggregation rules to improve the text. It turns out, however, that by formulating some rules in terms of discourse structure one can significantly reduce the complexity of the aggregation process. If aggregation is performed without discourse structure structure planning,

the aggregator has to inspect every pair of input elements for each aggregation rule it has, an order n^2 operation per rule for n elements, while if aggregation is performed after structuring, the aggregator need only inspect the pairs of elements within the discourse segments that directly contain the material to be generated, a reduction to (typically) two or three elements. In the example, the paragraph structure involves three parallel ELABORATION relations; see Figure 8 (a). In order to improve this text, the following three aggregation rules were applied:

1. *If two instances of the same RST relation emanate from a single nucleus, then merge the two instances into one relation, and merge their satellites into the same leaf node.*

2. *If several instances of the same RST relation appear in a* LIST, *then promote the relation, and* LIST *the respective nuclei and satellites together.*

3. *If input elements A and B within the same leaf node of the discourse structure contain the same action, the same ending date or time, and the same location, and they contain different actors, then merge the elements.*

The result generated was:

```
MEKAR-87 takes place in the South China Sea from 10/20 until 11/13.
Knox, Fanning, and Whipple are participating.  Knox and Fanning
arrive on 10/20.  Whipple arrives on 10/29.  Knox leaves on 10/31.
Fanning and Whipple leave on 11/13.
```

Of course, the general problem of aggregation for fluent text involves many non-structural issues as well (see for example [Dale 88, Van Dijk & Kintsch 83, Hovy 87]). But having access to the discourse structure enables one to begin addressing this problem in a realistic way. For some recent work see [Horacek 92].

3.3.7 Relations and Text Formatting

This problem deals with the formatting of written text[6]. Little written discourse — certainly no conference papers, reports, talk slides, etc. — is written completely without headings, section titles, occasional italicized portions, etc.; and much discourse contains itemized lists, footnotes, indented quotations, boldfaced terms, and other formatting devices.

Why? The reason is clear: each such formatting device carries a distinct meaning, and writers select the device that best serves their communicative intent at each point in the text.

A more interesting question is: How? That is, how do writers know what device to use at each point? How is device selection integrated with the discourse production process in general? Can

[6]This work was done by the author and Dr. Yigal Arens of USC/ISI.

```
                    (a)                                   (b)
                     |                                     |
                   ELAB                                    |
                   /  \                                  ELAB
                  /    \                                 /  \
              ELAB      \                               /    \
              /  \       \                      PT-WHOLE      LIST
             /    \       \                       / \        / | \
         ELAB      \       \                     /   \      /  |  \
         /  \       \       \               mekar  part SEQ  SEQ   SEQ
        /    \       \       \                          / \  / \   / \
  PT-WHOLE   SEQ     SEQ     SEQ                       Ka Kl Fa Fl Wa Wl
    /  \     / \     / \     / \
mekar part  Ka Kl   Fa Fl   Wa Wl
```

```
              (c)                                  (d)
               |                                    |
               |                                    |
             ELAB                                 ELAB
             /  \                                 /  \
            /    \                               /    \
     PT-WHOLE    SEQ                      PT-WHOLE     SEQ
       / \       / \                        / \       /   \
      /   \     /   \                      /   \     /     \
  mekar  part LIST   LIST              mekar  part LIST     LIST
              / | \  / | \                         /  \     /  \
             Ka Fa Wa Kl Fl Wl                 Ka,Fa  Wa Kl  Fl,Wl
```

Figure 8: (a) Original paragraph structure. (b) After rule 1: merging same relations. (c) After rule 2: merging relations in lists. (d) After rule 3: merging noun phrases.

the two processes be automated — can a text production system be made to plan not only the content and structure of the text but also the appropriate textual formatting for it?

The answer is yes, and this section describes an experiment that demonstrates this ability.

Textual Devices

In the course of our work on automated modality selection in multimedia communication [Hovy & Arens 90, Arens & Hovy 90a], we noticed an interesting fact: not only are the different text layouts and styles (plain text, itemized lists, enumerations, italicized text, inserts, etc., which we call here *Textual Devices*) used systematically in order to convey information, but *it is possible to define their communicative semantics precisely enough for them to be used in a text planner.* What's more, the systematicity holds across various types of texts, genres, and registers of formality. It is found in books, articles, advertisements, papers, letters, and even memos. The information these devices convey supplements the primary content of the text.

Though manuals of style (such as [CMS 82, APA 83, Van Leunen 79]) may seem relevant, they contain little more than precise descriptions of the preferred forms of textual devices in fact. We therefore classified textual devices into three broad classes — *Depiction*, *Position*, and *Composition* — and tried to provide functional descriptions of them. In all three cases, their communicative function is to delimit a portion of text for which certain exceptional conditions of interpretation hold. The following are some general uses of these devices:

- **1. Depiction:** selection of an appropriate letter string format.
  - *Parentheses:* text is tangential to the main text.
  - *Font switching:* text has special importance (new term, of central importance, foreign expression). *when the surrounding text is* not *italicized*).
  - *Capitalization:* text string names (identifies) an entity.
  - *Quotation marks:* text was written by another author.
- **2. Position:** Repositioning of text blocks.
  - *Inline:* non-distinguished normal case.
  - *Offset* (horizontal repositioning): text was authored by someone else.
  - *Separation* (vertical repositioning): text addresses a single point (a paragraph) or identifies subsequent text (headings or titles).
  - *Offpage:* text provides explanatory material (appendix, footnote).
- **3. Composition:** imposition of an internal structure on the text.
  - *Itemized list:* set of discourse objects on the same level of specificity with respect to the subject domain, each more than a clause (e.g., this list of devices).

- *Enumerated list:* set of discourse objects on the same level of specificity with respect to the domain, which are ordered along some underlying dimension, such as time, distance, importance.
- *Term definition:* pair of texts separated by a colon or other delimiter, in which the first names a discourse object and the second defines or explains it (e.g., this item in the itemized list).

Selecting appropriate textual devices relies on the author's ability to accurately characterize the meaning expressed by the specific portion of text as well as its relationship to the surrounding text (after all, the same sentence can properly be a footnote in one text and a parenthesized part of the text proper in another). Thus (ignoring such issues as textual prominence and style), the problem has three parts: the underlying semantic content to be communicated, the discourse structure, and the textual devices available. With respect to semantics, we took a standard approach (namely, using frame-like representation structures that contain terms from a well-specified ontology). To define the communicative semantics of textual devices, we employed an extension of RST.

Extending the Planner: An Example of Layout Planning

The RST text structure planner described in Section 3.2 was used for this experiment to plan and generate paragraphs of text about procedures to be followed by air traffic controllers, as represented for the ARIES system [Johnson & Harris 90, Johnson & Feather 91], an automatic programming project. In one experiment, the structurer was activated with the goal to describe the procedure to be followed by an air traffic controller when an aircraft is "handed over" from one region to the next. The underlying representation for this example consists of a semantic network of 18 instances, defined in terms of 27 air traffic domain concepts and 8 domain relations, implemented as frames in the Loom knowledge representation system [MacGregor 88]. The planner builds the paragraph tree shown in Figure 9.

Though the form of the text closely mirrors that of the actual Air Traffic Control Manual [ASA 89], the differences in formatting are significant; and these differences make the manual much more readable. The manual contains headings, term definitions signaled by italicized terms, enumerated lists, etc. After a study of several instructional texts, including recipes, school textbooks, and manuals for cars, sewing machines, and video players conducted at USC/ISI and the University of Nijmegen [Arens, Hovy, & Vossers 91, Vossers 91], we concluded that certain textual formatting devices are highly correlated with specific configurations of the underlying text structure tree. For example, a series of nested SEQUENCES, such as appears in Figure 9, is usually realized in the text as an enumerated list. Exceptions occur (in general) only when the individual items enumerated are single words (in which case the whole list is realized in a single sentence) or when there are few enough of them to place in a paragraph in-line (though usually in this case the keywords *first*, *second*, etc., are added).

```
              |
              |
             COND
            /    \
make-handoff   ELAB-PROCESS-STEP
                  /      \
        relay-info        SEQ
                         /   \
                   give-1     SEQ
                             /   \
                       give-2     give-3
```

```
When making a handoff, the transferring controller relays information
to the receiving controller in the following order. He gives the
target's position. He gives the aircraft's identification. He gives
the assigned altitude and appropriate restrictions.
```

Figure 9: Discourse structure for Air Traffic Control domain.

On the assumption that we can capture most of the reasons for using such formatting devices as enumerations on the basis of RST alone, we augmented the text plan SEQUENCE in order to include explicit text formatting commands and adapted the structure planner accordingly. For the formatting commands we used LaTeX forms such as `\begin{enumerate}` `\item` `\end{enumerate}` [Lamport 86]. Although our implementation was done within the framework of our specific generation technology, we believe a similar augmentation could be performed with most if not all the text planners being developed at this time. The resulting tree (with formatting commands indicated) is shown in Figure 10; the resulting text, generated by Penman and run through LaTeX, appears as:

> *When making a handoff, the transferring controller relays information to the receiving controller in the following order.*
>
> 1. *He gives the target's position.*
> 2. *He gives the aircraft's identification.*
> 3. *He gives the assigned altitude and appropriate restrictions.*

Semantics of Textual Devices

Despite its rather extreme simplicity, however, the example demonstrates that as long as one can characterize textual formatting devices in terms of configurations within the discourse structure, one can plan appropriate formatting commands of several types. The textual devices with structural definitions are:

```
                         |
                         |
                        COND
                      /      \
         make-handoff      ELAB-PROCSTEP
                              /      \
                   relay-info        SEQ-1
                                     /   \
("\begin{enumerate} \item" give-1)   (SEQ-2 "\end{enumerate}")
                                       /    \
                   ("\item" give-2)   ("\item" give-3)
```

Figure 10: Augmented discourse structure for Air Traffic Control domain.

- *Enumeration:* As described in the example above, the text structure relation SEQUENCE can generally be formatted as an enumerated list. The enumeration follows the sequence of the relation, which is planned in expression of some underlying semantic ordering of the items involved, for example time, location, etc.

- *Itemization:* The textual structure that relates a number of items without any underlying order is the RST relation LIST, which can be realized by an itemized list (unless the items are small enough to be placed into a single sentence).

- *Appendix, footnote, and parentheses:* These are three devices that realize the same textual relation, namely BACKGROUND. They differ in the amount of material included in the relation's Satellite.

- *Section title or heading:* This device realizes the textual relation IDENTIFICATION, which links an identifier with the body of material it heads. A section or subsection is appropriate when the IDENTIFICATION is combined with a SEQUENCE chain that governs the overall presentation of the text.

The insight that the communicative semantics of text formatting devices can to a significant extent be stated in terms of discourse structure relations is a powerful one. Two major limitations should however be borne in mind: additional factors determine the use of most formatting devices, and the representational power of current theories of discourse structure is still very limited. For some textual devices, no discourse relation has been identified by discourse linguists (for example, the Quotation device realizes the linguistic relation Projection, which is not included in the taxonomy in Figure 6 because it was not encountered in the survey). Other textual devices work on

a level too detailed for text coherence theories, since they operate on individual words within a clause. And finally, for some textual devices no purely linguistic constructs exist to handle them either (devices such as italicization and capitalization for word definition or emphasis cannot at this time be represented).

However, despite the problems with definitional delicacy, one can use discourse structure relations to define many of the textual devices listed above. To this extent, the incorporation of discourse structure relations into text planners is a new and very useful capability.

3.4 A New Architecture for Text Planning

All the work described in the previous sections lead up to a single conclusion: a new text planner had to be built to incorporate the more sophisticated definitions of intersegment relations, theme and focus control, intention, etc. This planner would require a simple basic architecture and a clean, open design, to facilitate the inclusion of all the disparate types of knowledge and the coding of their interrelationships.

This section[7] describes the new text planner that is being built jointly at USC/ISI and at GMD-IPSI. It is based on theoretical studies and experiments in text coherence (e.g., Rhetorical Structure Theory [Mann & Thompson 88], Conjunctive Relations [Martin 92]), theories of discourse (e.g., [Grosz & Sidner 86, Polanyi 88]), and text planning (e.g., [Hovy 88, Moore & Paris 89, Moore 89]), significantly advancing on those ideas and handling several new aspects of the problem.

This new text planner was designed to address several problems that we had encountered in the text planning work mentioned in previous sections and had observed in other, similar enterprises. An important motivation was a clearer separation of declarative and procedural knowledge in a generation system, as well as the identification of the distinct types of knowledge necessary to generate a text. It had become clear from a study of the current systems that as the planners' plan libraries grew, the same information (e.g., requirements of use and other preconditions) had to be represented several times, and it became harder to add still more plans and to modify existing plans because of their interrelationships. Also, existing planning systems often mixed information regarding the planning process and information necessary for linguistic realization in one single plan operator. Furthermore, some of the linguistic knowledge necessary to plan a text was often encoded in the planner itself, rendering the process more opaque. To address these problems, the new design was to make as clear as possible the distinction between procedural and declarative

[7]This research was jointly performed with the text planning group at USC/ISI, which included Mr. Giuseppe Carenini (IRST Institute, Italy), Mr. Thanasis Daradoumis (University of Barcelona, Spain), Dr. Julia Lavid (University of Madrid, Spain), Ms. Elisabeth Maier (IPSI, Germany), Mr. Vibhu Mittal (USC), Dr. Cécile Paris (USC/ISI), and Mr. Richard Whitney (USC/ISI), as well as the author. Portions of this section of the document were written by Maier, Lavid, Paris, and Mittal as well.

information, and to identify precisely and separate out the different types of knowledge required for creating a discourse structure.

3.4.1 Knowledge Resources Required for a Text Planner

The text planner embodies an attempt to isolate and use some of the major knowledge resources required to plan multisentential text. This section presents the major knowledge resources that we have so far identified, namely: *text types, communicative goals, schemas, discourse structure relations*, and, finally, a resource to handle *theme development and focus shift.*

In some cases, the knowledge resources actually represent the order of some planning operations. Such resources were implemented as systemic networks; they are the discourse relations and theme patterns. In other cases, the knowledge resources provide information which the planner uses to make decisions. Such resources were implemented as property-inheritance networks; they are the text types, communicative goals, and schemas. Both types of representation are declarative, enabling one to capture inherent commonalities within the resource, and promote notational clarity and simplicity of processing.

Each node in either type of network may contain one or more *realization operators* which indicate the effects of choosing the node, such as making additions to the discourse structure, choosing subsequent nodes to visit, setting requirements upon subsequent grammatical realization, etc. (for a full list see Section 3.4.2). Knowledge resources co-constrain each other via these realization operators. Section 3.4.2 describes how the property-inheritance networks are used and the systemic networks are traversed during the planning process, and how a text structure is built during the traversal.

This planner is far from complete. Motivations for various choices have not been fully identified and several important text planning functions, such as noun phrase planning, lexical choice, lexical cohesion, and sentence structure planning, are lacking altogether. These problems are briefly discussed in Section 3.4.4.

Text Type Hierarchy

It has long been observed that certain types of linguistic phenomena (e.g., the rhetorical structure, lexical types, grammatical features) closely reflect the genre of the text (e.g., scientific papers, financial reports). A text generation system that contains a rich set of expressive possibilities requires some representation of genres or text types in order to constrain its options, since no other resource will provide the necessary information, and the system will be unable to choose between alternative formulations.

Several text typologies have been proposed by linguists. To mention only a few: [Biber 89] identified 8 basic types of texts, based on statistically derived grammatical and lexical commonal-

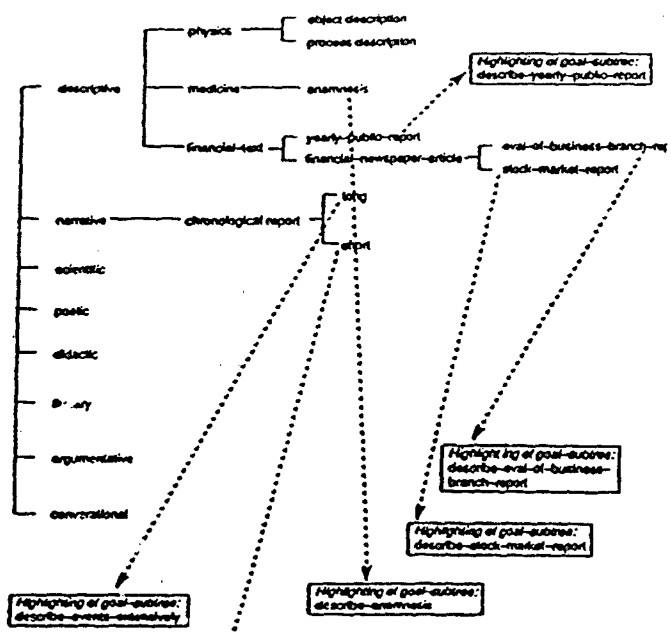


Figure 11: Hierarchy of text types.

ities. The Washington School has proposed a detailed classification of different genres of written scientific and technical English [Trimble 85], additionally pointing out typical relationships within and between rhetorical/textual units. [De Beaugrande 80] proposed a general classification of text types, also arguing that text types determine the types of discourse structure relations used.

Given its generality, De Beaugrande's hierarchy of text types was selected as a basis for the text planner's text types, with extensions as needed to handle text types particular to the domains addressed. The hierarchy (partially shown in Figure 11) is represented as a property-inheritance network in the knowledge representation system Loom [MacGregor 88]. Each text type in this hierarchy has associated with it the constraints it imposes on other resources, such as which communicative goals it entails, which discourse relations it favors, any appropriate grammatical constraints, etc. As a result, once a type has been established for the text to be generated, the selection of other parameters used during the generation process can be constrained appropriately (for instance, interpersonal discourse relations almost never appear in objective scientific reports, while love letters tend to contain mainly those relations). Thus the planner's predefined text types help pre-select or de-activate certain options in the generation process.

Communicative Goal Hierarchy

As have been used in many generation systems, communicative goals describe the discourse

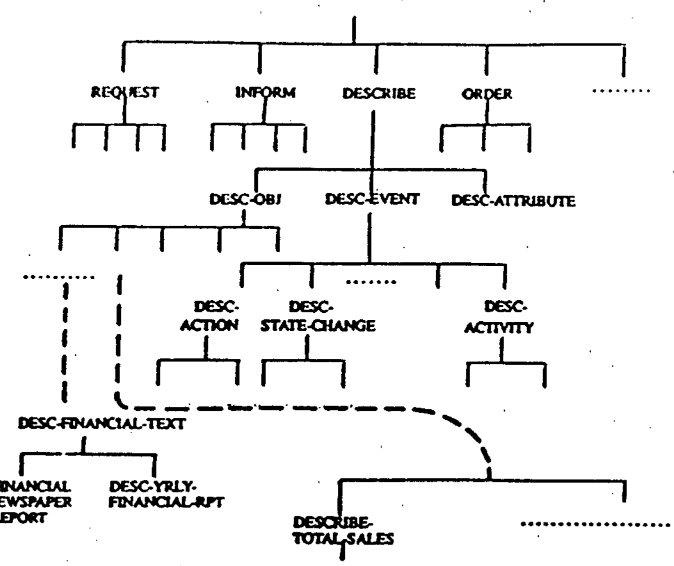


Figure 12: Hierarchy of communicative goals.

purpose(s) of the speaker. The planner contains a rudimentary taxonomization of communicative goals, starting at the topmost level with some very general goals, such as INFORM, DESCRIBE, REQUEST, and ORDER, which are eventually refined into specific goals to describe (or relate, etc.) specific types of information for specific contexts (see Figure 12). The taxonomy, which is implemented as a property-inheritance network, resembles the one being derived from Speech Acts by Allen and his colleagues; see [Allen 91]).

Each discourse segment (a subtree of the discourse structure) is headed by one of these goals as its discourse segment purpose, and schema stages and discourse structure relations can contain goals as well. Each communicative goal contains one or more realization operators — instructions for the planner to perform specific actions (see Section 3.4.2). The planner's lowest clause-level goals are called *planner primitive speech acts*; these goals apply at the leaves of the discourse structure and signal that the next step is grammatical realization.

Schemas

In many circumstances, texts exhibit a stereotypical structure. In text planning systems, such structure is usually represented in schemas which specify the topics of discussion that appear in the text as well as their ordering (see Section 3.3.4). The stages of structural stereotypes can be defined at the clausal level (indicating the type of process of each sentence to be included and its position), but can equally well be defined at a more general level (indicating the sequence of general topics to be included). Linguists have proposed several schema-like approaches to model such structure: e.g., macrostructures [Van Dijk & Kintsch 83], holistic structures [Mann & Thompson 88], and the

Generic Structure Potential [Halliday & Hasan 85]. Recognizing the utility of such structures, we include them (represented within a property-inheritance network) into the planner[8].

As an example, a schema to generate financial reports could contain the following communicative goals in the dictated order: (1) describe-total-sales-briefly (heading); (2) describe-total-sales; (3) describe-domestic-sales; (4) describe-export-sales and (5) describe-future-outlook. Section 3.4.3 describes how this schema is used by the planner to generate a particular text.

Just as the previous two resources co-constrain the other resources (e.g., the choice of text type can influence the selection of a schema), the instantiation of a schema can highlight or suppress different discourse relations, or the various stages of a schema can favor particular theme development patterns.

Discourse Structure Relations

Many linguists and computational linguists have studied the relationships that hold between sentences or segments of text, identifying relations that they claim need to hold in order for a text to be coherent (e.g., [Grimes 75, Mann 84, Hobbs 78, Mann & Thompson 88, Sanders et al. 91, Redeker 90]. These relations must be used in a generation system in order to guide the selection and organization of the information to be included when other structuring guidance is lacking, such as when a schema stage calls for more material than can fit into a single clause. The necessity and use of discourse structure relations in text planners to ensure coherence has been amply discussed (e.g., [Hovy 88, Moore & Paris 89, Paris 90, Cawsey 90, Maybury 90]).

The new planner contains three networks of discourse relations, implemented as systemic networks. The networks were based on several main sources: the relations defined in Rhetorical Structure Theory [Mann & Thompson 88], which were extended in Hovy's taxonomization of a collection of the relations proposed by over 30 researchers from various fields (later reorganized with Maier; see [Hovy 90b, Maier & Hovy 91], and Section 3.3.3), and Martin's linguistically inspired taxonomization of the conjunctive relations [Martin 92]. The relations were divided into three major portions, corresponding to the three major functions of language (semantic/ideational, interpersonal, and pre-presentational/textual); portions of the networks appear in Figure 13 and Figure 14. When organizing material, the planner is free in the general case to establish several discourse relations (typically, one for each of the major functions) between the existing discourse structure and the new piece of material; as shown in the networks, the selection of ideational, interpersonal, and textual relations is not exclusive. As with the other resources, the discourse relation networks co-constrain the other knowledge resources, by for example preselecting theme patterns or specifying aspects of grammatical realization.

[8] In spite of the frozen nature of schemas, the underlying rhetorical relationships among the different parts of each schema still exist. Given sufficient knowledge, a system should be able to plan out the same text without using a schema. However, lacking a complete specification of all the resources required in generation, a planner can use schemas as a useful source of 'compiled knowledge' and so avoid the need to re-derive structures over and over again.

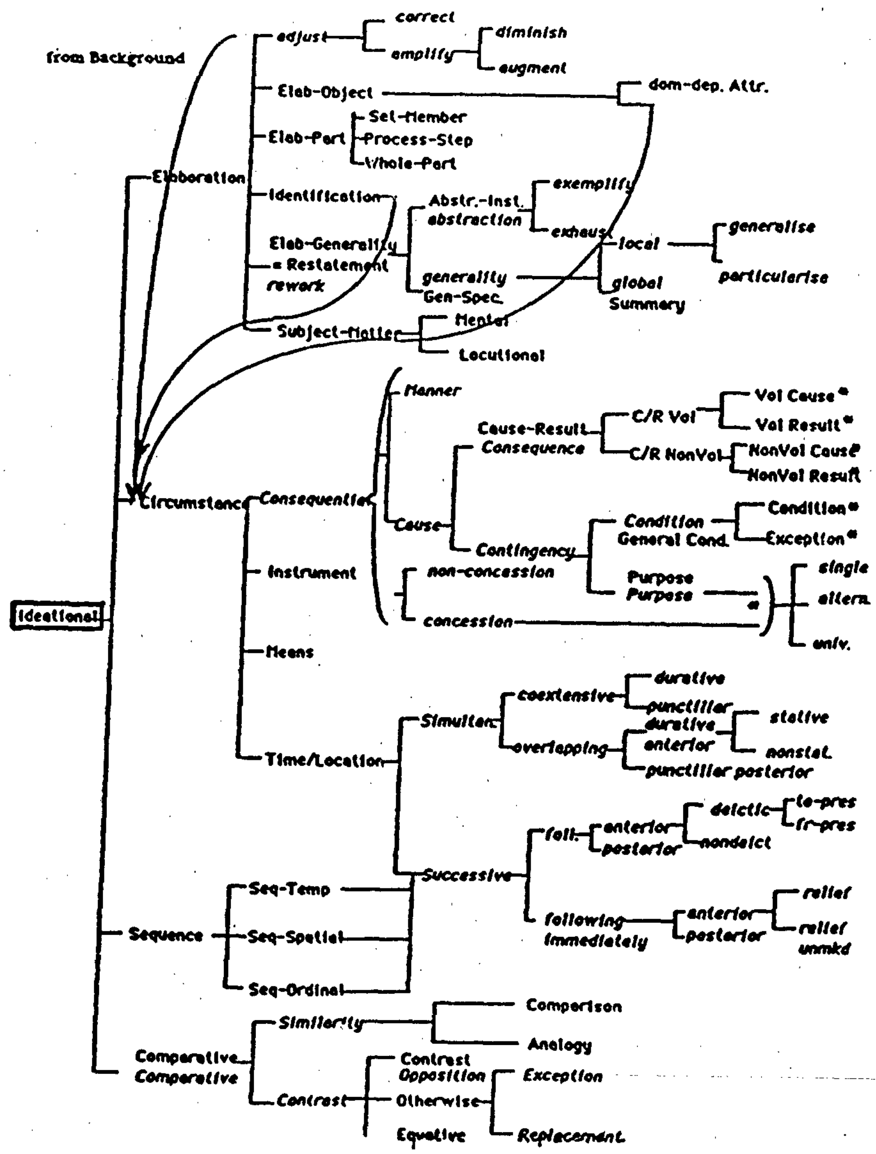


Figure 13: Discourse structure relations: ideational network.

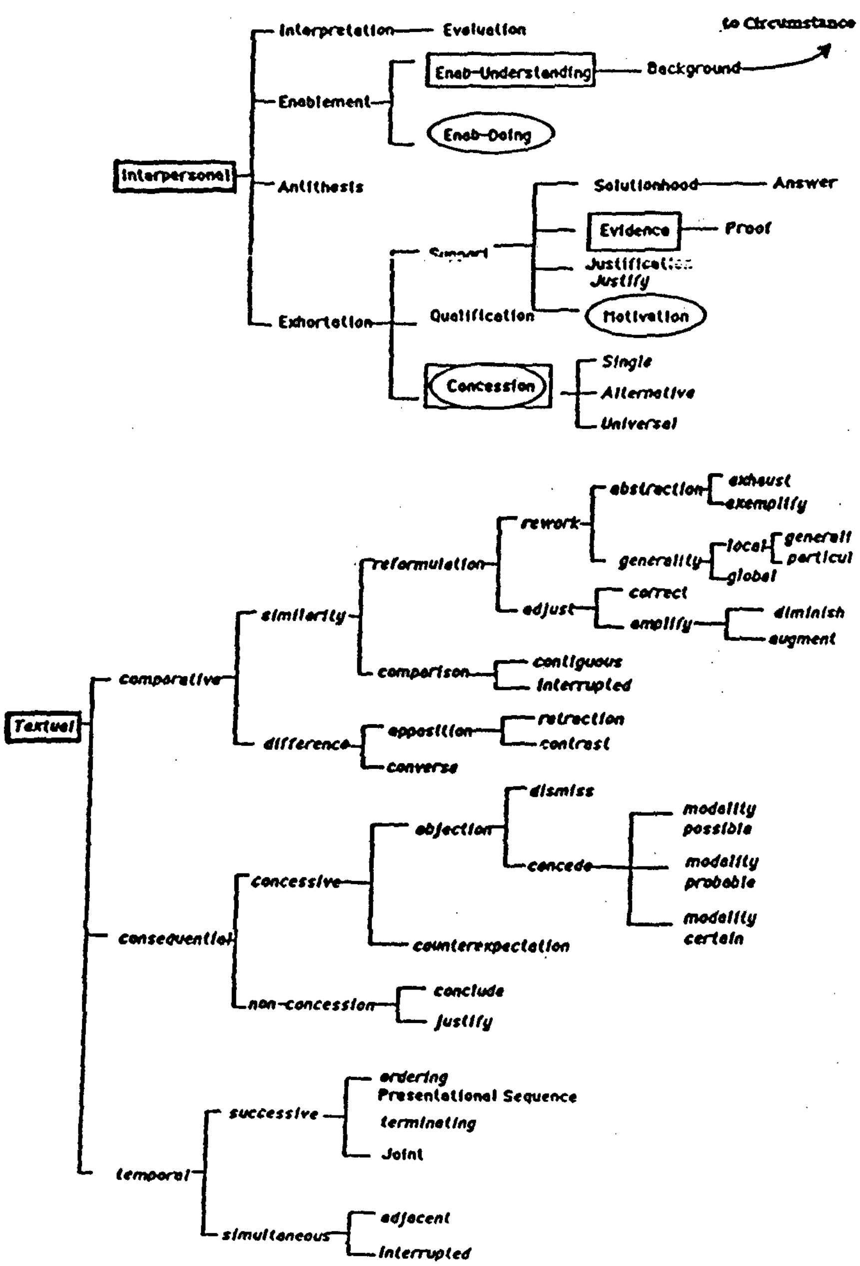


Figure 14: Discourse structure relations: interpersonal and textual networks.

Theme Development Information

Careful linguistic and computational studies have shown the need for a resource describing the potential theme developments and shifts of focus (see for example [Halliday 85, Quirk et al. 72]) in order to signal the introduction of a new topic of discussion and to provide its thematic relationship to previous topics. These concerns have not been the subject of much computational work (but see [Sidner 83, McCoy & Cheng 88]); in text generation they have taken the form of so-called focus shift rules (see [McKeown 85, McCoy 85, Paris 91, Hovy & McCoy 89], and Section 3.3.5). Unfortunately, these rules have usually been implemented procedurally and with little regard to the true complexity of the issues underlying them. In the new text planner the potentialities of theme development are represented declaratively in a systemic network (see Figure 15).

Though the study of theme has been traditionally been restricted to the sentence level, it also plays a role at the the clause-complex and even discourse levels. This should be taken into consideration by a text generation system. Given a text to be generated, the system must establish how theme development may proceed and how themes are to be marked in each clause. The following three concerns arise:

- the type of theme to select: following Halliday (85) , there can be three different and simultaneous themes in each clause: the ideational (or topical; expressing processes, participants, or circumstances), the interpersonal (expressing modal meanings such as probability, usuality, or opinion), and the textual (such as continuatives — "yes,", "well,", "oh,", or conjunctions). The first type is semantically required.
- the theme progression pattern involved: the new theme can be the same as the theme of the previous clause; it may be part of the rheme of the previous clause; or it may be an element of what is called the "hypertheme" or general discourse segment topic (see [Daneš 74]): note also the similarity to the focus shift rules of Sidner and McKeown).
- the linguistic degree of markedness of the theme: realization depends on the type of clause.

The motivations behind each choice follow pragmatic principles of information processing, including:

- *the Topic-Comment constraint* [Werth 84, Giora 88], also known as the *Graded Informativeness requirement*: a message is maximally effective if information which is presumed or given in the context is presented before information which is new;
- *the Processibility principle* [Leech 83]: a text should be constructed so that it is easy to process in real time, by placing the focus tone group at the end of the clause (the *maxim of end-focus*) and the "heavy" constituents in final position (the *maxim of end-weight*);
- *discourse relation requirements* [Mann & Thompson 88]: some discourse relations have a canonical (unmarked) order of surface realization.

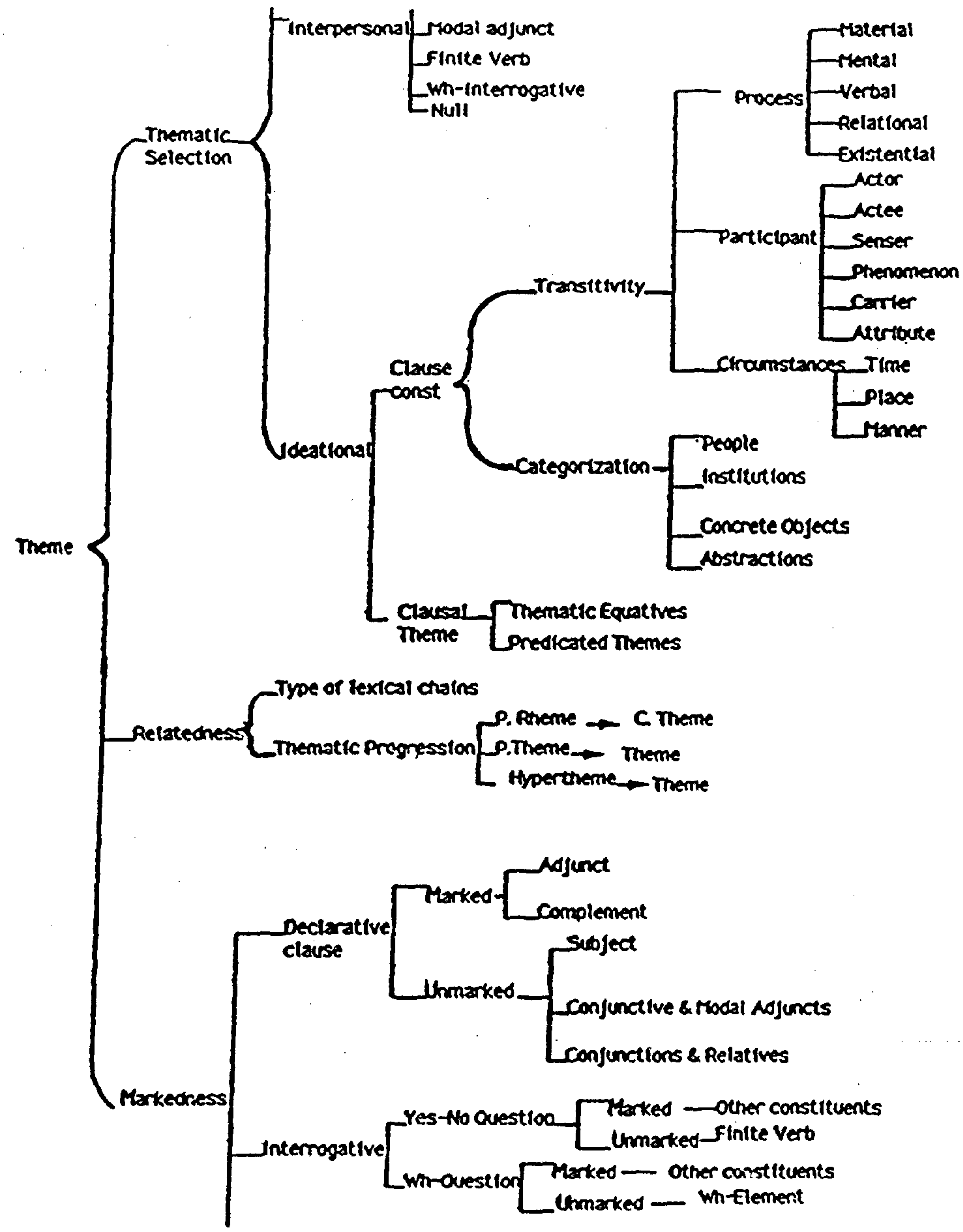


Figure 15: A portion of the theme network.

3.4.2 The Planning Process

Planning with the networks proceeds analogously to the generation of single sentences with Penman [Penman 89, Mann 83, Mann & Matthiessen 85, Hovy 90c]: in both cases, the traversal mechanism proceeds through the network, causing traversal choices to be made at nodes (systems, and building a tree-like structure as a result. We implemented the network in Penman's internal notation so as to be able to reuse some of its traversal code.

Associated with each node in the networks is an inquiry function which queries the environment in order to determine which branch to follow, and a set of realization operators that instruct the planner what to do next.

The planning operation is very simple. After an initial setup phase, the system simply executes a basic planning cycle over and over again until planning is complete. In the setup phase, the user activates the planner with a communicative goal, as described in Section 3.4.1, which causes the selection of a desired text type, and is then posted on the goal stack and simultaneously on the Discourse Structure Tree. Then the basic planning cycle begins. Essentially, this cycle proceeds as follows: First, the planner checks whether there is a realization on the agenda. If so, it performs the realization by applying its action to its parameters. If there are no realizations left, the planner checks whether there is a discourse goal on the goal-stack. If there is, the planner finds the realizations associated with the goal and loads them onto the agenda; if no discourse goals remain, the planning is done.

Clearly, the action of the system lies in the realizations. Each realization is an instruction to be performed. At present, the system uses the following realizations:

1. `(ACTIVATE-SCHEMA schema-name)`: Find the schema and load its realizations onto the agenda.
2. `(ADD-TO-D-STRUC goal concept parentpos)`: Add the given communicative goal into the discourse structure tree at the given position.
3. `(CHANGE-HYPERTHEME -chainofroles-)`: Change the topic under discussion to the filler of the given chain of roles, starting from the current topic.
4. `(HIGHLIGHT-COMM-GOALS -goals-)`: Highlight the given goals so that only they will be considered for future planning.
5. `(HIGHLIGHT-RELATION -relations-)`: Start traversal of the discourse relations network(s) at the given relations, using the current topic of discussion.
6. `(BLOCK-RELATION -relations-)`: Mark the given discourse structure relations so that they cannot be traversed for the remainder of the current sentence.

7. (PREFER-THEME conceptrole): Add instructions for the realization component that the given role of the topic under discussion should be thematized in the clause.

8. (SET-MACROTHEME concept): Change the overall topic of discussion.

9. (SET-UP-DISCOURSE-GOAL goal): Activate the given goal: load it onto the goal stack and into the discourse structure tree at the current growth point and add its realizations to the agenda.

10. (TRAV-ONE-NETWORK-NODE node-name): Locate the given node in the knowledge resource networks, apply its inquiry function, record the result (the inquiry choice), and load the realizations associated with the result onto the agenda.

3.4.3 An Example of the Planner in Action

This section provides a brief trace in order to show how the various linguistic resources interact to guide the construction of the discourse structure. The example is a text from a bank's annual report:

> Declines in Total Sales of the Swiss Cheese Union
>
> (1) In the business year 1986/87 (ending July 31), the 40 cheese trading firms associated in the Swiss Cheese Union sold 79,035 tons of cheese altogether, equal to a 2.6% decline. (2) Domestic sales of table cheeses enjoyed a relatively positive trend, with Swiss households buying 22,100 tons of their preferred cheeses, a gain of 3.9% from one year earlier.
>
> (3) Exports benefited from brisk demand in the early months of the year and since inventories continued to register normal volumes, export prices could be raised by about 5% at the beginning of 1987. (4) But a few months later the incoming order volume levelled off again with the consequence that export volumes narrowed by 4.3% to 47,100 tons for the 12 months of the business year. (5) Exports of Greyerzer (7.4%) were hardest hit by the drop, whereas the decline in the case of Emmentaler was more moderate at 4.2%. (6) Sales in Italy, the leading market for Emmentaler, gained in line with an advertising campaign which had been launched in the closing months of the past business year and recovered to last year's level. (7) Export losses were most extensive in the case of shipments to France, the United States, Spain and Belgium. (8) In contrast to this, more Emmentaler cheese was marketed than in the previous business year in the Federal Republic of Germany, the United Kingdom and Canada. (9) Exports of Sbrinz recorded a surprisingly favorable trend with a gain of 4.2%.
>
> (10) The outlook for the sale of Swiss table cheeses must be assessed with reserve in view of the stiff competition. (11) Little scope remains in the domestic or export business for quantitative or pricing improvements.

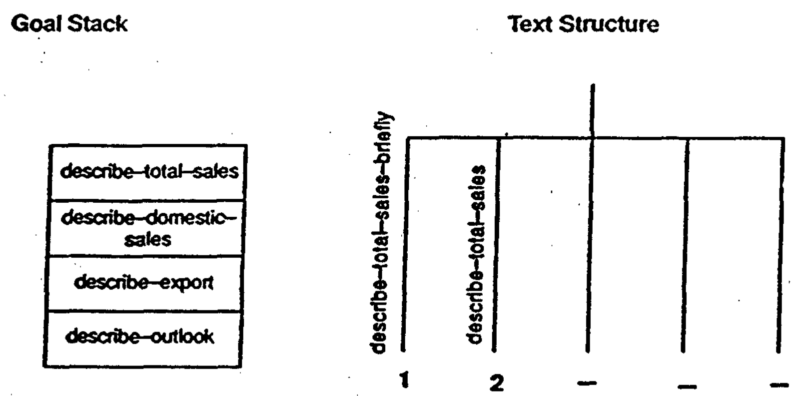


Figure 16: Snapshot of the text planner state.

The semantic information in this text was represented in the Loom knowledge representation system [MacGregor 88].

Given GENERATE-YEARLY-PUBLIC-REPORT as communicative goal and CHEESE-UNION-SALES-86 as topic of discussion, the schema mentioned in Section 3.4.1 is activated, and the planner goes through the stages indicated in the schema. Let us assume now that the first two clauses — the headline and the first proposition — have already been generated. The state of the discourse structure and the text appears in Figure 16.

After generating the first two clauses, the next active goal (the goal on the top of the goal stack) is DESCRIBE-TOTAL-SALES. The planner activates this goal by popping it off the stack, loading it onto the discourse structure at the current point of growth, and then checking its definition in the goal hierarchy for any realization statements to be performed. In this case, there is only one: highlight the discourse structure relation `interpretation`. This realization is loaded onto the agenda. This completes the planning cycle.

The next cycle begins. The planner checks the agenda and finds the just-loaded realization. It performs the realization by highlighting `interpretation` in the interpersonal relations network, which causes the planner to check whether any topic material with that relation to the current topic of discussion can be brought into the discourse. This check is performed by an inquiry function that accesses the planner environment with a question that can be paraphrased as:

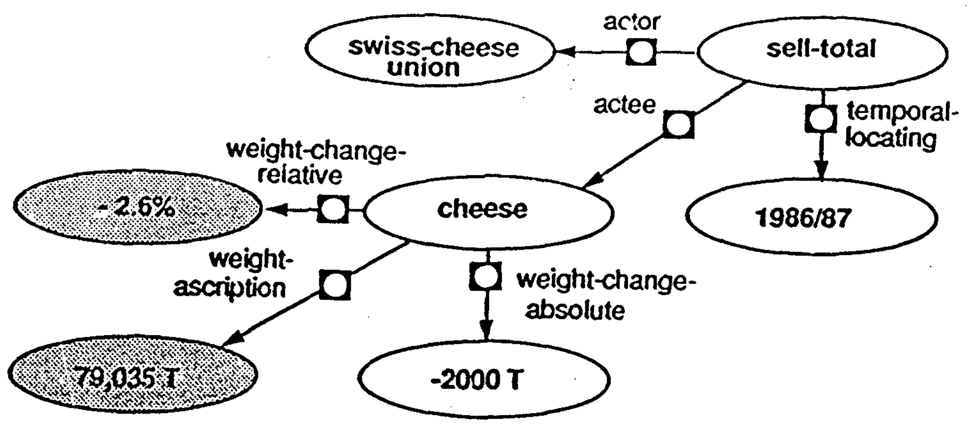


Figure 17: Fragment of the Domain Model.

Interpretation-Q-Code:
"Was a numerical value mentioned in the last proposition and can it be expressed in relation to other values?"

From the information about the topic (as contained in the knowledge representation system), a possible candidate for such a relation is the value of the role `weight-ascription`. The inquiry code retrieves a role and a value which fulfills the above condition: the role `weight-change-relative` represents the weight ascription relative to that of the preceding year. The relevant segment of the domain model appears in Figure 17.

The successful finding of this material signals the applicability of the relation `interpretation`. The planner thus activates the realization statements associated with this relation, in this case:

- **knowledge selection:** Each relation contains specifications of the material it relates. The realization associated with `interpretation` selects both the absolute and the relative ascriptions for the weight, linked with the concept `corresponds`, and calls for the building a new instance of the relation accordingly.
- **discourse structure growth:** This realization calls for the addition of the new instance of the `interpretation` relation at the current growth point in the discourse structure.
- **theme determination:** This realization calls for traversal of the theme network in order to determine the thematization pattern of the new clause or clauses.

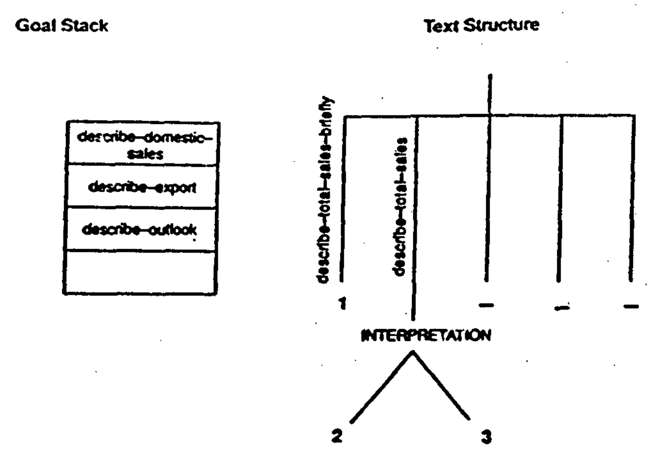


Figure 18: Discourse structure after the new relation has been planned.

- **operations on relations:** To prevent the repetitive use of the interpretation relation (which would lead to a monotonous text), this realization calls for `interpretation` to be blocked for further use until the end of the next sentence.

The planner loads these four realizations onto the agenda and thereby completes its cycle.

In the next cycle, the planner runs the knowledge selection realization listed above and builds the new relation. In the following cycle it adds the relation to the discourse structure. And so forth; the resulting form of the discourse structure after these realizations appears in Figure 18.

Space considerations prevent a detailed description of the remaining planning. In essence, the planning cycle keeps repeating, first handling all the realizations on the agenda and then all the goals on the goal stack, until no more remain.

3.4.4 Conclusion

This section briefly described the architecture and functioning of the new text planner currently being developed jointly at USC/ISI and GMD-IPSI. It is based on the idea that the linguistic resources needed to generate coherent text (as well as their interrelationships) should be represented explicitly, separately, and distinct from the procedural knowledge required for text planning. The planner is described in more detail in [Hovy et al. 92].

There is no claim that all the knowledge sources required to produce coherent discourse have been identified. The problems of lexical choice, the planning of noun groups (and referring expressions in general), the problem of sentence delimitation are all unaddressed in the planner. In

addition, planning of lexical cohesion has also been left out[9]. We do, however, believe that the architecture of our planner lends itself well to the incorporation of additional knowledge resources when they become available. The representational power of systemic networks — interlocking options that capture the potentialities of expression — and the clear and simple planning cycle offer, we hope, sufficient scaffolding for the needs of text planning of the future.

3.5 Generalization of the Work to Multiple Media

During the course of the research described above, it became increasingly clear that text planning can be viewed as a special case of a more general kind of communicative planning, namely, planning communication within a multimodal environment. The two problems share many aspects, and their solutions seem to lie so closely together that the development of a joint solution seems a natural path to take[10].

When communicating, people almost always employ multiple modalities. No single medium seems to suffice; for example, natural language, which is after all the most powerful representational medium developed by humankind, is still usually augmented by pictures, diagrams, etc. (when written) or by gestures, hand and eye movements, intonational variations, etc. (when spoken). We are investigating the knowledge people use and the processes by which they use it to produce multimedia communications and to interpret them. In particular, we ask: How do people apportion the information to be presented to various modalities? And how do they reassemble the portions into a single message again?

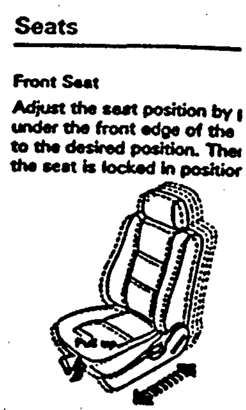
From our work in multimedia human-computer interactions [Vossers 91, Hovy & Arens 91, Hovy & Arens 90, Arens & Hovy 90a, Arens & Hovy 90b], we have come to appreciate the complexity of the task of mustering all the communication resources and orchestrating them to contribute to the intended message in a coherent way. Our work is an effort to construct a fairly detailed set of representational terms that capture all the factors that play a role in multimedia communication. It includes an extensive survey of relevant literature from Psychology, Human-Computer Interfaces, Natural Language Processing, Linguistics, Human Factors, and Cognitive Science (see [Vossers 91]). Our preliminary analysis of the knowledge required just to support bimodal communication (we limited ourselves to language and diagrams only) has uncovered well over a hundred distinct factors that play a role in the higher level aspects of the production and

[9] The idea of cohesion as a unity-creating device is well-known in linguistics [Halliday & Hasan 76, Ventola 87] and has recently been discussed also in the A.I. literature (see [Morris & Hirst 91]). The study of lexical cohesion is not only interesting because it determines how well constructed a text is, but also because the patterns of cohesion reveal something about the semiotic organization of texts, that is, about the way a text is realized in stages. One of our first priorities when extending the present planner will be to make this resource operative in the form of a network.

[10] This research was performed together with Dr. Yigal Arens from USC/ISI, who was funded by a grant from DARPA.

interpretation processes, as well as over fifty rules that express the interdependencies among these factors.

In this work, we have discovered an unexpected and somewhat satisfying result with cognitive import: many of the rules that express the interdependencies between relevant factors operate cross-modally; that is to say, *the same rule can be used to control the parsing or generation of some aspect of both a diagram and a piece of text.* We believe that the parsimony and expressive power of these rules simultaneously motivates the particular representation level we have used for the factors and also suggests how the complex task of multimedia communication is achieved with less cognitive overhead than at first seemed necessary.



To make this clearer, we present a small example immediately. In the diagram on the right (taken from a Honda car manual; for a fuller discussion of this example see below), our analysis led us to the following result. On analyzing the heading **Seats**, we identified a collection of presentational features (including *boldface, large-font*, etc.) that differed quite substantially from the features describing the label **Pull up**. However, the communicative functions of the two items turned out to be closely related types of *naming*, and hence fulfilled very similar author goals. Going back to the presentation, we found that the presentational features causing the difference were in fact superficial ones, ones that served merely to ensure the differentiation of the item against its presentational background. This superficial difference could be captured in a single rule about *distinguishing*, thereby enabling the very different items (a text heading and a diagram label) to be handled with the same rule. This example is discussed in more detail in the next section.

The example, when described, seems obvious. But it can only be explained by using such notions as *distinguished/separated* (both the positional/off-text distinctiveness and the realizational/text-vs-graphics distinctiveness) and *communicative function* (one part of the communication serves to name/introduce/identify another part). When one constructs a vocabulary of terms on this level of description, one finds unexpected overlaps in communicative functionality across modalities. These overlaps can be exploited to reduce the rules required to parse and generate multimedia communications. The implications for human communication lie in the significant simplification of an extremely complex task: the production and interpretation of communications in multiple modalities.

3.5.1 A Framework that Supports Multimedia Communication

What factors play a role in multimedia communication? In particular, how does a producer determine which information to allocate to which modality, and how does a perceiver segment the

communication into parts, recognize the function of each part, and integrate the separate functions into a coherent whole?

In trying to answer these questions, we took instruction manuals, and we limited ourselves, for now, to just two modalities: natural language text and line diagrams. We first studied the knowledge required to perform multimedia communication — the static factors that play a role — and have recently started studying the processing involved — the dynamic activities that make use of that knowledge to generate and understand actual presentations. We addressed the knowledge by dividing the problem into four parts, believing that multimedia communication is influenced by factors from:

- the intentions, desires, and characteristics of the producer,
- capabilities of the perceiver,
- the nature of the information to be conveyed, and
- the characteristics of the media used.

For each of these aspects separately, we followed a three-step methodology: first, we identified the phenomena that seem to play a role (e.g., the fact that the producer often wants to affect the receiver's future goals, or the fact that different media utilize fewer or more 'dimensions'); second, we characterized the variability involved in each phenomenon (e.g., a producer may want to affect the receiver's goals through warnings, suggestions, hints, requests, etc., or language is expressed in a 'linear' fashion way while diagrams are two-dimensional); and third, we mapped out the interdependencies among all the values of all the phenomena. The results are networks of interdependencies in which each node represents a single phenomenon and each arc a possible value for it, and the arcs are joined and split by AND and OR connectors into an AND/OR network to express the interdependencies (this network form is used extensively by Systemic-Functional linguists to represent grammars of various languages in exactly the same way; see [Halliday 85]).

Although we have not yet implemented our results in a working system, it is our intention to do so following closely the work of the Penman project [Penman 89, Mann & Matthiessen 85, Hovy 90c]. In this project, the grammar of English is represented as an AND/OR network of the form described above and sentence generation proceeds by traversing the network from 'more semantic' toward 'more syntactic' nodes, collecting at each node features that instruct the system how to build the eventual sentence (see [Matthiessen 84]). Parsing proceeds by traversing the same network 'backwards', eventually arriving at the 'more semantic' nodes and their associated features, the set of which constitutes the parse and determines the parse tree (see [Kasper & Hovy 90, Kasper 89]).

This bi-directionality of processing is one advantage of the network representation form. Another is its independence of process issues; one can implement the knowledge we have distilled in a

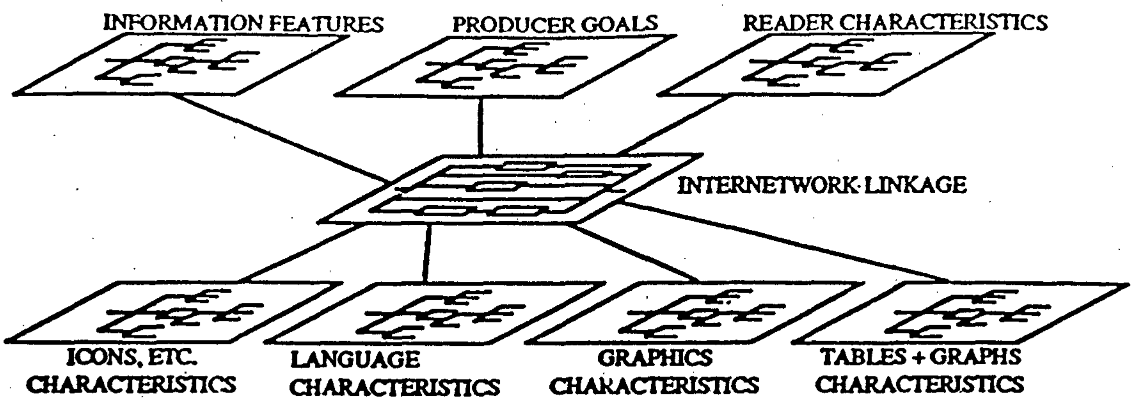


Figure 19: Framework of Knowledge Resources to Support Multimedia Communication.

traditional rule-based system as well as in a connectionist one. Of course, many process issues still have to be faced, but those questions we will address later.

The overall design appears in Figure 19. Each knowledge resource appears as a separate network; the central network houses the interlinkages between the other ones. When producing a communication, the communicative goals and situation cause appropriate features of the upper three networks to be selected, and information then propagates through the interlinkage network (the system's 'rules') to the appropriate modality networks at the bottom, causing appropriate values to be set, which are used in turn to control the low-level generation modules (the language generator, the diagram constructor, etc.). When analyzing a communication, appropriate features in the relevant bottom networks are selected for each portion of the communication, and the information is propagated upward to select appropriate 'high-level' features that describe the producer's goals, the information for that portion, etc.

The next section provides more details about the individual knowledge resources and illustrates the interlinking rules with examples.

3.5.2 Knowledge Resources for Multimedia Communication

The information presented in this section is derived from an analysis of pages from instruction manuals for appliances (such as user manuals for a motor car, a sewing machine, a VCR, as well as a cookbook) and from readings in the related literature. In the networks, curly brackets mean AND (that is, when entering one, all paths should be followed in parallel) and square brackets EXCLUSIVE OR (that is, at most one path must be selected and followed). Square brackets with slanted serifs are INCLUSIVE OR (that is, zero or more paths may be selected and followed). Whenever a feature is

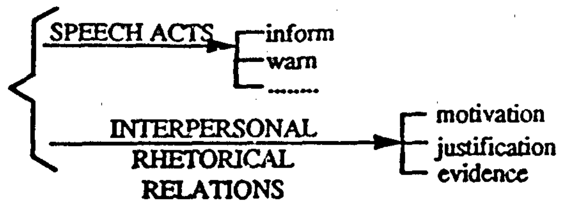


Figure 20: Portion of the Producer Goals Network.

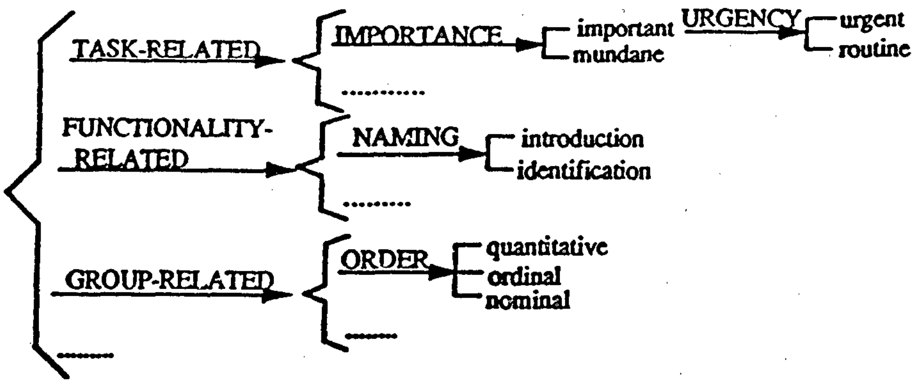


Figure 21: Portion of the Information Features Network.

passed during traversal of the network, it is to be recorded, for it co-determines the eventual result. We can provide here only very small fragments of some of the networks; for more information, see [Vossers 91].

Figure 20 provides a portion of the network containing the aspects of a producer's communicative intentions that may affect the appearance of the communication. In this network fragment *warn* is distinguished from *inform* because, unlike inform speech acts, the semantics of warnings involve capturing the attention of the reader in order to affect his/her goals or actions. To achieve this, a warning must be realized using presentation features that distinguish it from the background presentation. The mechanism for achieving this presentation is described later (Figure 23).

Figure 21 provides a portion of the network describing the features of information that affect its display. Some of those are:

- **Importance**
  - *Important*: The information relates to the user's persistent goals (involving actions which could cause personal injury or property damage). *Important* information must be reinforced by textual devices, such as 'boldface', 'capitalization', etc., to give text a notable

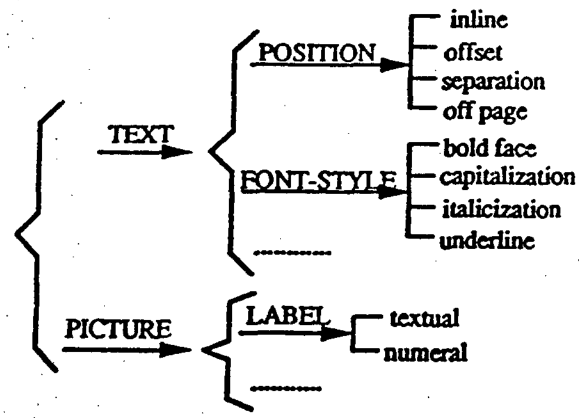


Figure 22: Portion of the Modalities Network.

appearance [Hovy & Arens 91].

- *Mundane*: The normal, non-distinguished case.

- **Naming**
  - *Identification*: The information identifies a portion of the presentation. An *Identification* relation may exist between, e.g., a text-label and a picture part.
  - *Introduction*: The information identifies and introduces other information. An *Introduction* relation exists between a text heading and the following material.
- **Order**
  - *Quantitative*: The items of a conceptually and/or syntactically parallel set of information items may be ordered by the value of some measure they express. E.g., temperature readings for various days.
  - *Ordinal*: The items of a set of information items may be ordered according to the semantics of events they describe. E.g., steps in a cooking recipe.
  - *Nominal*: The items are not inherently ordered.

Figure 22 provides a portion of the network describing the characteristics of the modalities that determine the form of the presentation of information and hence constrain their use. The terms used in the network are self-explanatory.

The central component of the knowledge used in processing multimedia presentations is the collection of rules which establish associations between goals of the producer and the content of the information (Figures 20 and 21 respectively), and the surface features of presentations (Figure 22).

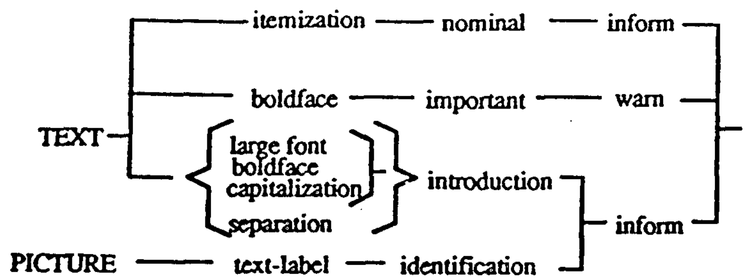


Figure 23: Portion of the Internetwork Linkage.

A small portion of these rules, also represented in network form, appears in Figure 23. Moving from left to right through the network, one first finds the presentation forms which express the information, then features of the information which are linked to various presentation forms, and finally the producer goals. The use of the various rules captured in this network is illustrated next.

3.5.3 Examples

This section contains examples of the use of the previously described knowledge in multimedia communication. The domain is the page explaining how to adjust the front seat of the Honda Accord [Honda Manual]; see Figure 24.

Example 1:
Refer to the section heading **Front Seat** and the label **Pull up** in Figure 24. The section heading is analyzed as having features *text-in-text*, *boldface*, *separation*, *short*. The label is analyzed as having features *text-in-picture*, *short*.

On first inspection, the section heading **Front Seat** and the label **Pull up** look very different. But after following the internetwork linkage rules in Figure 23, both items are seen to serve related producer goals; *introduce* and *identify*, respectively. These are both instances of *naming* (see Figure 21). The features that differ are simply those that cause each item to be distinguished against its background. The operative rule appears to be:

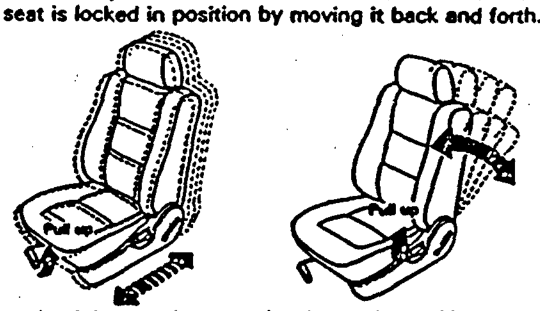
> To indicate *naming*, use *short* text which is *distinct* from the background presentation object.

Within a picture, *distinction* is achieved by the mere use of text. Within text, however, *distinction* must be achieved by varying the features of the surrounding rendering of the language. Features varied may be the font type and size, or the position of the fragment in question in relation to the general flow.

Seats

Front Seat

Adjust the seat position by pulling up on the adjustment lever under the front edge of the seat cushion, and sliding the seat to the desired position. Then release the lever and make sure the seat is locked in position by moving it back and forth.



The angle of the seat-back can be changed by pulling up on the lever at the outside edge of the seat and leaning forward or back. Release the lever and let the seat lock into place when it reaches the desired angle.
Seat belts are most effective when the driver and passenger are sitting up straight and well back in each seat.

⚠WARNING

- Do not adjust the driver's seat while the car is moving; the seat could move suddenly, causing loss of vehicle control.
- To reduce the risk of sliding under the seat belts in a collision, the seat-backs should be reclined no more than is necessary for comfort.

Figure 24: Page from Honda manual.

Now this notion of *distinction* turns out to be usefui also for completely different applications.

Example 2:

Consider the text bullets at the bottom of the figure. Its function is to *warn* and not just *inform*, which is the purpose of the preceding paragraphs. The text in question has the feature *bold*, but we see again that this serves to *distinguish* the warning text from the background, thus signaling the special force of *warn*. One can predict that to warn a reader concerning information displayed in a diagram or picture, it will suffice to place text within the non-textual substrate.

The notion of *distinction* did not explicitly exist in the networks — Figure 23 indicates it with an appropriate collection of specific features. Its importance was discerned in the course of investigating the internetwork linkage rules and their application to presentations such as this manual page.

3.5.4 The Complexity of the Problem

A survey of the literature on the design of presentations (book design, graphic illustration, etc.; see [Bretin 83, Tufte 83, Tufte 90]) underscores how much this area of communication is an art and how hard it is to describe the rules that govern presentations. But people clearly do follow rules when they use complementary modalities in their communication; not any random text paragraph of a book, for example, can be illustrated with a diagram.

Psychologists have for years been studying multimedia issues such as the effects of pictures in text, design principles for multimedia presentation, etc. [Dwyer 78, Fleming & Levie 78, Hartley 85, Twyman 85]. However, most of these results are too general to be directly applicable in work that is to be computationalized.

On the other hand, cognitive science studies of the past few years have provided useful results which should be incorporated into theories about good multimedia design [Petre & Green 90, Larkin & Simon 86, Mayer 89, Roth & Mattis 90]. They address questions such as whether graphical notation is really superior to text, what makes a picture worth (sometimes) more than a thousand words, how illustration affects thinking, the characterization of data etc.

Recent work in the area of multimedia interfaces is a promising beginning toward a more formal and computational theory. [Mackinlay 86] described the automatic generation of a variety of tables and charts; [Feiner 88, Wahlster et al. 91, Arens et al. 88, Neal 90] illustrate various aspects of the processing and knowledge required for automated multimedia computer presentations. But there is still a long way to go; all of these systems barely scratch the surface of the general problems involved in reasoning about multimedia presentations.

The next section describes the work done under this contract to address some aspects of this problem.

3.6 Toward Multimodal Presentation Planning

While we do not pretend to have a theory to explain the phenomena, we believe that a careful study of the types of modalities people use, and the types of information they typically utilize them for, will single out characteristics of the underlying cognitive representations and shed light on people's communicative processes. With these issues in mind, initiating a study of the characteristics of representation as expressed through communication, we decided to examine first two aspects:

- communication-related characteristics of information
- modes of human-human and human-computer communication

In addition, it is clearly necessary to take into account the modes of interaction with computers as well, in order eventually to test the rules developed and implemented on a computer against the display decisions made by people. A vocabulary must be developed to identify the characteristics salient to the display of information. This vocabulary should:

- describe all features of the information that are salient for presentation purposes,
- describe all features of presentation modalities that can be utilized to convey information,
- be general enough to allow comparisons and specific enough to differentiate between different modalities and information.

3.6.1 Characterization of Modalities

The following terms are used to describe presentation-related concepts. We take the point of view of the communicator (indicating where the consumer's subjective experience may differ).

1. Consumer: A person interpreting a communication.

2. Modality: A single mechanism by which to express information. Examples: spoken and written natural language, diagrams, sketches, graphs, tables, pictures.

3. Exhibit: A complex exhibit is a collection, or composition, of several simple exhibits. A simple exhibit is what is produced by one invocation of one modality. Examples of simple exhibits are a paragraph of text, a diagram, a computer beep. Simple exhibits involve the placement of one or more **Information Carriers** on a background **Substrate.**

4. Substrate: The background to a simple exhibit. That which establishes, to the consumer, physical or temporal location, and often the semantic context, within which new information is presented to the information consumer. The new information will often derive its meaning, at least in part, from its relation to the substrate. Examples: a piece of paper or screen (on which information may be drawn or presented); a grid (on which a marker might indicate the position of

an entity); a page of text (on which certain words may be emphasized in red); a noun phrase (to which a prepositional phrase may be appended). An empty substrate is possible.

5. Information Carrier: That part of the simple exhibit which, to the consumer, communicates the principal piece of information requested or relevant in the current communicative context. Examples: a marker on a map substrate; a prepositional phrase within a sentence predicate substrate. A degenerate carrier is one which cannot be distinguished from its background (in the discussion below the degenerate carrier is a special case, but we do not bother explicitly to except it where necessary. Please assume it excepted).

6. Carried Item: That piece of information represented by the carrier; the 'denotation' of the carrier.

For purposes of rigor, it is important to note that a substrate is simply one or more information carrier(s) superimposed. This is because the substrate carries information as well[11]. In addition, in many cases the substrate provides an internal system of semantics which may be utilized by the carrier to convey information. Thus, despite its name, not all information is transmitted by the carrier itself alone; its positioning (temporal or spatial) in relation to the substrate may encode information as well. This is discussed further below.

7. Channel: An independent dimension of variation of a particular information carrier in a particular substrate. The total number of channels gives the total number of independent pieces of information the carrier can convey. For example, a single mark or icon can convey information by its shape, color, and position and orientation in relation to a background map. The number and nature of the channels depend on the type of the carrier and on the exhibit's substrate.

3.6.2 Internal Semantic Systems

Some information carriers exhibit an internal structure that can be assigned a 'real-world' denotation, enabling them subsequently to be used as substrates against which other carriers can acquire information by virtue of being interpreted within the substrate. For example, a map used to describe a region of the world possesses an internal structure — points on it correspond to points in the region it charts. When used as a background for a ship icon, one may indicate the location

[11]Note that from the information consumer's point of view, Carrier and Substrate are subjective terms; two people looking at the same exhibit can interpret its components as carrier and substrate in different ways, depending on what they already know. For example, different people may interpret a graph tracking the daily value of some index differently as follows: someone who is familiar with the history of the index may call only the last point of the graph, that is, its most recent addition, the information carrier, and call all the rest of the graph the substrate. Someone who is unfamiliar with the history of the index may interpret the whole line plotted out as the information carrier, and the graph's axes and title, etc., as substrate. Someone who is completely unfamiliar with the index may interpret the whole graph, including its title and axis titles, as information carrier, and interpret the screen on which it is displayed as substrate.

of the ship in the world by placing its icon in the corresponding location on the map substrate. Examples of such carriers and their internal semantic systems are:

Carrier	Internal Semantic System
Picture	'real-world' spatial location based on picture denotation
NL sentence	'real-world' sentence denotation
Table	categorization according to row and column
Graph	coordinate values on graph axes
Map	'real-world' spatial location based on map denotation
Ordered list	ordinal sequentiality

Other information carriers exhibit no internal structure. Examples: icon, computer beep, and unordered list.

An internal semantic system of the type described is always intrinsic to the item carried.

3.6.3 Characteristics of Modalities

In addition to the internal semantics listed above, modalities differ in a number of other ways which can be exploited by a presenter to communicate effectively and efficiently. The values of these characteristics for various modalities are shown in Table 1.

Carrier Dimension: Values: *0D, 1D, 2D.* A measure of the number of dimensions usually required to exhibit the information presented by the modality.

Internal Semantic Dimension: Values: *0D, 1D, 2D, >2D, 3D, #D, ∞D.* The number of dimensions present in the internal semantic system of the carrier or substrate.

Temporal Endurance: Values: *permanent, transient.* An indication whether the created exhibit varies during the lifetime of the presentation.

Granularity: Values: *continuous, discrete.* An indication of whether arbitrarily small variations along any dimension of presentation have meaning in the denotation or not.

Medium Type: Values: *aural, visual.* What type of medium is necessary for presenting the created exhibit.

Default Detectability: Values: *low, medlow, medhigh, high.* A default measure of how intrusive to the consumer the exhibit created by the modality will be.

Baggage: Values: *low, high.* A gross measure of the amount of extra information a consumer must process in order to become familiar enough with the substrate to correctly interpret a carrier on it.

Generic Modality	Carrier Dimension	Int. Semantic Dim.	Temporal Endurance	Granularity	Medium Type	Default Detectability	Baggage
Beep	0D		transient	N/A	aural	high	
Icon	0D		permanent	N/A	visual	low	
Map	2D	2D	permanent	continuous	visual	low	high
Picture	2D	3D	permanent	continuous	visual	low	high
Table	2D	2D	permanent	discrete	visual	low	high
Form	2D	>2D	permanent	discrete	visual	low	high
Graph	2D	2D	permanent	continuous	visual	low	high
Ordered list	1D	#D	permanent	discrete	visual	low	low
Unordered list	0D	#D	permanent	N/A	visual	low	low
Written sentence	1D	∞D	permanent	discrete	visual	low	low
Spoken sentence	1D	∞D	transient	discrete	aural	medhigh	low
Animated material	2D	3D	transient	continuous	visual	high	high
Music	1D	?	transient	continuous	aural	med	low

Table 1: Modality characteristics.

3.6.4 How Carriers Convey Information

As part of an exhibit, a carrier can convey information along one or more **channels**. For example, with an icon carrier, one may convey information by the icon's shape, color, and possibly through its position in relation to a background map. The number and nature of the channels depends on the type of carrier and the substrate.

The semantics of a channel may be *derived* from the carrier's spatial or temporal relation to a substrate which possesses an internal semantic structure; e.g., placement on a map of a carrier representing an object which exists in the charted area. Otherwise we say the channels is *free*.

Among *free* channels we distinguish between those whose interpretation is *independent* of the carried item (e.g., color, if the carrier does not represent an object for which color is relevant); and those whose interpretation is *dependent* on the carried item (e.g., shape, if the carrier represents an object which has some shape).

Most of the carrier channels can be made to vary their presented value in time. Time variation

can be seen as an additional channel which provides yet another degree of freedom of presentation to most of the other channels. The most basic variation is the alternation between two states, in other words, a flip-flop, because this guarantees the continued (though intermittent) presentation of the original basic channel value.

3.6.5 Characterization of Information and Its Presentation

In this section we develop a vocabulary of presentation-related characteristics of information.

Broadly speaking, as shown in Table 2, three subcases must be considered when choosing a presentation for an item of information: intrinsic properties of the specific item; properties associated with the class to which the item belongs; and properties of the collection of items that will eventually be presented, and of which the current item is a member. These characteristics are explained in the remainder of this section.

Type	Characteristic	Values
Intrinsic Property	Dimensionality	0D, 1D, 2D, >2D, ∞D
	Transience	live, dead
	Urgency	urgent, routine
Class Property	Order	ordered, nominal
	Density	dense, discrete, N/A
Set Property	Volume	singular, little, much

Table 2: Information characteristics by type.

Dimensionality: Some single items of information, such as a data base record, can be decomposed as a vector of simple components; others, such as a photograph, have a complex internal structure which is not decomposable. We define the *dimensionality* of the latter as *complex*, and of the former as the dimension of the vector.

Since all the information must be represented in some fashion, the following must hold (where *simple* dimensionality has a value of 0, *single* the value 1, and so on, and *complex* the value ∞):

The Basic Dimensionality Rule of Presentations

Dim(Info) ≤ Dim(Carrier) + Free Channels(Carrier) + Internal Semantic Dim(Substrate)

In addition, we have found that different rules apply to information of differing dimensions. With respect to dimensionality, we divide information into four classes as follows:

- *Simple:* Simple atomic items of information, such as an indication of the presence or absence of email.
  - As carrier, use a modality with a dimension value of 0D.
  - No special restrictions on substrate.
- *Single:* The value of some meter such as the amount of gasoline left. Associated rule is:
  - No special restrictions on substrate.
- *Double:* Pairs of information components, such as coordinates (graphs, map locations), or domain-range pairs in relations (automobile $\times$ satisfaction rating, etc.).
  - As substrate, use modalities with internal semantic dimension of 2D.
  - As substrate, use modalities with discrete granularity (e.g., forms and tables) if information-class of both components is discrete.
  - As substrate, use modalities with continuous granularity (e.g., graphs and maps) if information-class of either component is dense.
  - As carrier, use a modality with a dimension value of 0D.
- *Multiple:* More complex information structures of higher dimension, such as home addresses. It is assumed that information of this type requires more time to consume (hence the last rule in this group).
  - As substrate, use modalities with discrete granularity if information-class of all components is discrete.
  - As substrate, use modalities with continuous granularity if the information-class of some component is dense.
  - As carrier, use a modality with a dimension value of at least 1D.
  - As substrate and carrier, do not use modalities with the temporal endurance value transient.
- *Complex:* Information with internal structure that is not decomposable, such as photographs.
  - Check for the existence of specialized modalities for this class of information.

Transience: Transience refers to whether the information to be presented expresses some current (and presumably changing) state or not. Presentations differ according to:

- *Live:* The information presented consists of a single conceptual item of information (that is, one carried item) that varies with time (or in general, along some linear, ordered, dimension), and for which the history of values is not important. Examples are the amount of money owed while pumping gasoline or the load average on a computer. Most appropriate for *live* information is a single exhibit.
  - As carrier, use a modality with the temporal endurance characteristic transient if the update rate is comparable to the lifetime of the carrier signal.
  - As carrier, use a modality with the temporal endurance characteristic permanent if update rate is much longer.
  - As substrate, unless the information is already part of an existing exhibit, use the neutral substrate.
- *Dead:* The other case, in which information does not reflect some current state, or in which it does but the history of values is important. An example is the history of some stock on the stock market; though only the current price may be important to a trader, the history of the stock is of import to the buyer.
  - As carrier, use ones that are marked with the value permanent temporal endurance.

Urgency: Some information may be designated *urgent*, requiring presentation in such a way that the consumer's attention is drawn. This characteristic takes the values *urgent* and *routine*:

- *Urgent:* This situation is exemplified in emergencies, whether they be imminent meltdowns or a warning to a person crossing the road in front of a car. Rules of modality allocation are:
  - If the information is not yet part of a presentation instance, use a modality whose default detectability has the value high (such as an aural modality) either for the substrate or the carrier.
  - If the information is already displayed as part of a presentation instance, use the present modality but switch one or more of its channels from fixed to the corresponding temporally varying state (such as flashing, pulsating, or hopping).
- *Routine:* The normal case.
  - Choose a modality with low default detectability and a channel with no temporal variance.

Density: The difference between information that is presented equally well on a graph and a histogram and information that is not well presented on a histogram is a matter of the density of

the class to which the information belongs. The former case is *discrete* information; an example is the various types of car made in Japan. The latter is *dense* information; an example is the prices of cars made in Japan.

- *Dense:* A class in which arbitrary small variations along a dimension of interest carry meaning. Information in such a class is best presented by a modality that supports continuous change:
  - As substrate, use a modality with granularity characteristic continuous (e.g., graphs, maps, animations).
- *Discrete:* A class in which there exists a lower limit to variations on the dimension of interest. Appropriate modalities are as follows:
  - As substrate, use a modality with granularity characteristic discrete (e.g., tables, histograms, lists).

Volume: A batch of information may contain various amounts of information to be presented. If it is a single fact, we call it *singular*; if more than one fact but still little relative to some some task- and user-specific threshold, we call it *little*; and if not, we call it *much*. This distinction is useful because not all modalities are suited to present *much* information.

- *Much:* The relatively permanent modalities such as written text or graphics leave a trace to which the consumer can refer if he or she gets lost doing the task or forgets, while transient modalities such as spoken sentences and beeps do not. Thus the former should be preferred in this case.
  - As carrier, do not use a modality the temporal endurance value transient.
  - As substrate, do not use a modality the temporal endurance value transient.
- *Little:* There is no need to avoid the more transient modalities when the amount of information to present is *little*.
- *Singular:* A single atomic item of information. A transient modality can be used. However, one should not overwhelm the consumer with irrelevant information. For example, to display information about a single ship, one need not draw a map.
  - As substrate, if possible use a modality whose internal semantic system has *low* baggage.

3.6.6 An Example

We present three simple tasks in parallel.

	Coordinates	Name	Photograph
Information	48N 2E	Paris	Eiffel Tower
Dimensionality	*double*	*single*	*single*
Volume	*little*	*singular*	*singular*
Density	*dense*	*discrete*	*discrete*
Transience	*dead*	*dead*	*dead*
Urgency	*routine*	*routine*	*routine*

Table 3: Example information characteristics.

Given: the task of presenting Paris (as the destination of a flight, say).
Available information (three separate examples): the coordinates of the city, the name *Paris*, and a photograph of the Eiffel Tower.
Available modalities: maps, spoken and written language, pictures, tables, graphs, ordered lists.

The modality characteristics are listed among those in Table 1. The information characteristics are listed in Table 3.

The allocation algorithm classifies information characteristics with respect to characteristics of modalities, according to the rules outlined in Section 3.6.5. The modality with the most desired characteristics is then chosen to form the exhibit.

Handling the coordinates: As given by the rules mentioned in Section3.6.5, information with a *dimensionality* value of *double* is best presented in a substrate with a *dimension* value of *2D*. This means that candidate substrates for the exhibit are maps, pictures, tables, and graphs. Since the *volume* is *little*, *transient* modalities are not ruled out. The value *dense* for the characteristic *density* rules out tables. The values for *transience* and *urgency* have no further effect. This leaves tables, maps, and graphs as possible modalities. Next, taking into account the rules dealing with the internal semantics of modalities, immediately everything but maps are ruled out (maps' internal semantics denote spatial locations, which matches up with the denotation of the coordinates). If no other information is present, a map modality is selected to display the location of Paris.

Handling the name: The name Paris, being an atomic entity, has the value *single* for the *dimensionality* characteristic. By the appropriate rule (see Section 3.6.5), the substrate should be the neutral substrate or natural language and the carrier one with *dimension* of *0D*. Since the *volume* is *singular*, a *transient* modality is not ruled out. None of the other characteristics have any effect, leaving the possibility of communicating the single word **Paris** or of speaking or writing a sentence such as *"The destination is Paris"*.

Handling the photograph: The photograph has a *dimensionality* value *complex*, for which appropriate rules specify modalities with *internal semantic dimension* of *3D*, and with *density* of

dense (see Section 3.6.5) — animation or pictures. Since no other characteristic plays a role, the photograph can simply be presented.

3.6.7 Conclusion

The enormous numbers of possibilities made available when one attempts to deal with multiple modalities, as illustrated by the psychological, cognitive science, and automatic-generation work listed above, is daunting. While we hope that the modality-based analysis and knowledge representation work described here will contribute to a systematic understanding of the question, we take heart at the fact that many of the rules centrally involved in the information-to-modality linkage are capable of handling several modalities and several types of meaning. As we illustrated in this paper, the overlaps in communicative functionality of aspects from quite different modalities — for example, the spatial offset and distinct typefont of a heading and the different nature of a text label in a diagram both serve to identify and name the accompanying material, and can both therefore be handled by the same rule — suggests that the problem may be feasible for computational treatment after all. This somewhat surprising result may help explain why multimedia communication is so pervasive in human interaction.

4 Conclusion

The past three years have seen a significant new developments in several aspects of the task of language generation in human-computer interactions:

- increased understanding of the structure and intentional import of discourse as a phenomenon of language (see Section 3.1);
- increased knowledge and ability of Natural Language Processing specialists to perform the planning of paragraphs by computer (see Sections 3.2 and 3.4);
- creation of a general-purpose taxonomy of discourse structure relations culled from numerous sources (see Section 3.3.3);
- demonstration of feasibility of automated text format planning in tandem with text structure planning (see Section 3.3.7);
- increased knowledge about the underlying knowledge required to perform information-to-medium allocation in multimodal human-computer interactions (see Section 3.6).

Before 1987, the only general method available for generating multisentence text was the instantiation of so-called schemas, which, being essentially paragraph-sized templates, are limited

in flexibility and applicability. The new developments in text planning using RST and similar relation/plans, piloted at ISI and explored in several directions under this contract, and taken further in several respects by a number of other investigators over the past two years, for example by [Moore & Paris 89, Moore 89, Rankin 89, De Souza et al. 89, Maybury 90, Cawsey 90, Dobeš & Novak 91], promise well for our abilities to plan and generate longer, multiparagraph texts, well before the end of the decade. The new text planner resulting from this research, as developed at USC/ISI and IPSI and described in Section 3.4, points the way toward the kind of architecture that is simultaneously flexible and extensible enough to handle the demands of different domains and communicative intentions, rich enough to incorporate all the various types of information that play a role in the selection and planning of multisentence discourse as distinct knowledge resources, and open and clear enough to support the coding of the complex interdependencies that exist between them.

The work reported here is under continued development. The new text planner is being extended and used in the EXPECT project at USC/ISI; aspects of it may also be incorporated in the PANGLOSS Machine Translation project; the multimedia investigations are being continued by graduate students from the University of Nijmegen and USC. Additional funding will be sought to continue building upon the foundation already established. The eventual goal is to incorporate all knowledge resources — syntactic and semantic (from Penman), discoursal (from the new text planner), and multimedia — into a single framework to be used as a basis for further research in human-computer interactions. It is the intention to distribute the new text planner and, when ready, the new multimedia presentation manager, as a research vehicle the same way Penman is currently being distributed to research institutions and universities around the world.

Coupled with the ability to perform discourse analysis using the same discourse representation and semantic formalisms, all located in an integrated multimedia display system whose planner performs not only the display outlay planning and information-to-medium allocation, but the paragraph planning and text formatting as well, these new developments in text planning are an exciting and highly productive area of research in Natural Language Processing.

5 Outreach and Dissemination

5.1 Personnel

The Penman project currently consists of the following full-time staff: Dr. John Bateman, Dr. Eduard Hovy (project leader) and Mr. Richard Whitney. Dr. Bateman spends a significant portion of each year at the IPSI Institute in Darmstadt, Germany, where he leads the project's sister project KOMET in developing German capabilities for Penman. Closely associated with the project at USC/ISI are Dr. William Swartout, Dr. Cécile Paris, and Dr. Yigal Arens.

The work described in this document was performed principally by two groups at USC/ISI: the text planning group and the multimedia presentations group. In addition to Dr. Hovy and Mr. Whitney, the former group contained from USC/ISI Dr. Cécile Paris and Mr. Vibhu Mittal, members of the EES project, while the latter group contained Dr. Yigal Arens from the SIMS project. In addition to the permanent personnel, the project enjoyed the comments and assistance of several short-term and longer-term visitors, including:

- Dr. Julia Lavid (University of Madrid, Spain; Oct. 1990 - Dec. 1991);
- Ms. Elisabeth Maier (IPSI Institute, Darmstadt, Germany; Jan. 1991 and Mar. 1991 - Aug. 1991);
- Mr. Giuseppe Carenini (IRST Institute, Trento, Italy; Dec. 1990 - Mar. 1991);
- Ms. Mira Vossers (University of Nijmegen, Nijmegen, The Netherlands; Nov. 1990 - Aug. 1991).
- Mr. Thanasis Daradoumis (University of Barcelona, Spain; June 1991 - Aug. 1991);

5.2 Collaborations

In addition to the medium- and longer-term visitors, numerous researchers investigating different aspects of text planning, discourse, and generation visited the groups during the two and a half year lifetime of the contract. Several ongoing research efforts in text planning have or have had direct collaborative connections with members of the group, including the text planning work in the LILOG project at IBM Stuttgart, Germany (Dr. HaJo Novak and colleagues); the new multilingual planner being built at the University of Ulm, Germany (Dr. Dietmar Rösner and Dr. Chris Mellish from the University of Edinburgh); the text planning work being done at the University of Waterloo, Canada (rof. Chrysanne DiMarco and students), and the continuing collaboration between Dr. Cécile Paris and Prof. Johanna Moore from the University of Pittsburgh, PA.

In addition, in order to promote increased development of various computational aspects of Systemic Linguistics, the Penman project entered into a multinational collaboration in which various partners would have different focuses of research, while using the Penman sentence generator as a common center. While not directly involving the text planning work funded by this contract, the collaboration added to the intellectual richness and scope of ideas. This collaboration involves:

- A group in the Linguistics Department of the University of Sydney, Australia
- The KOMET project at IPSI, Darmstadt, West Germany
- The Penman project at ISI, Los Angeles, USA

5.3 Publications

The group has an active publication record. Since 1988, the following papers and/or presentations were published or made (or accepted for later publication) on work funded or partially funded by this contract (full versions are available from the author):

- Recent Trends in Computational Research on Monologic Discourse Structure. Hovy, E.H. *Computational Intelligence*, February 1992 (to appear).
- A New Level of Natural Language Generation Technology: Capabilities and Possibilities. Hovy, E.H. *IEEE Expert*, April 1992 (to appear).
- Employing Knowledge Resources in a New Text Planning Architecture. Hovy, E.H., Lavid, J., Maier, E., Mittal, V., and Paris, C.L. In *Proceedings of the 6th International Workshop on Language Generation*, Trento, Italy, April 1992 (to appear).
- Parsimonious or Profligate: How Many and Which Discourse Structure Relations? Hovy, E.H. and Maier, E. Submitted to *Computational Intelligence*, 1992.
- The Use of Intersegment Relations in Discourse Generation. Hovy, E.H. Submitted to *Artificial Intelligence*, 1991.
- Organizing Discourse Structure Relations using Metafunctions. Hovy, E.H. and Maier, E. Submitted to volume edited by H. Horacek, Bielefeld, 1991.
- Automatic Generation of Formatted Text. Hovy, E.H. and Arens, Y. In *Proceedings of the 9th AAAI Conference*, Anaheim, CA, July 1991.
- From Interclausal Relations to Discourse Structure — A Long Way Behind, a Long Way Ahead. Hovy, E.H. Keynote presentation at the *3rd European Workshop on Text Generation*, Innsbruck, Austria, March 1991.
- A Metafunctionally Motivated Taxonomy for Discourse Structure Relations. Hovy, E.H. and Maier, E. In *Proceedings of the 3rd European Workshop on Language Generation*, Innsbruck, Austria, March 1991.
- Descrying the Knowledge Underlying the Processing of Multimedia Instruction Manuals. Hovy, E.H., Arens, Y. and Vossers, M. Unpublished manuscript, 1991.
- Categorizing the Knowledge Used in Multimedia Presententations. Hovy, E.H., Arens, Y. and Vossers, M. In *Proceedings of the AAAI Workshop on Intelligent Multimedia Interfaces*, AAAI-91, Anaheim, CA, July 1991.

- Text Layout as a Problem of Modality Selection. Hovy, E.H. and Arens, Y. In *Proceedings of the 5th Conference on Knowledge-Based Specification* (pp. 87-94), RADC Workshop, Syracuse, NY, Sept. 1990.

- Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. Hovy, E.H. Presented at the *5th International Workshop on Language Generation*, Pittsburgh, PA, July 1990.

- How to Describe What? Towards a Theory of Modality Utilization. Hovy, E.H. and Arens, Y. In *Proceedings of the 12th Cognitive Science Conference*, Cambridge, MA, Aug. 1990.

- Explanation Generation in Historical Context: Two Methodologies of Investigating Discourse Structure. Hovy, E.H. Presented at the *AAAI Workshop on Comparative Analysis of Explanation Planning Architectures*, AAAI-91, Anaheim, CA, July 1991.

- Approaches to the Planning of Coherent Text. Hovy, E.H. In *Natural Language in Artificial Intelligence and Computational Linguistics*, Paris, C.L., Swartout, W.R. and Mann, W.C. (eds), Kluwer Publishers, Boston, 1990. Presented at the *4th International Workshop on Language Generation*, Santa Catalina Island, CA, July 1988. Also available as USC/Information Sciences Institute Research Report ISI/RS-89-245.

- Unresolved Issues in Paragraph Planning. Hovy, E.H. In *Current Research in Natural Language Generation*, Dale, R., Mellish, C., and Zock, M. (eds) (pp. 17-45). New York, NY: Academic Press, 1990. Presented at the *2nd European Workshop on Natural Language Generation*, Edinburgh, Scotland, April 1989.

- When is a Picture Worth a Thousand Words? — Allocation of Modalities in Multimedia Communication. Hovy, E.H. and Arens, Y. Presented at the *AAAI Spring Symposium on Human-Computer Interactions*, Palo Alto, CA, March 1990. Long version of paper available as unpublished document, ISI/USC.

- Focusing your RST: A Step toward Generating Coherent Multisentential Text. Hovy, E.H. and McCoy, K.F. In *Proceedings of the 11th Cognitive Science Conference*, Ann Arbor, MI, Aug. 1989.

- Notes on Dialogue Management and Text Planning in the LILOG Project. Hovy, E.H. Unpublished manuscript, LILOG, IBM Deutschland, Stuttgart, Germany, 1989.

- Planning Coherent Multisentential Text. Hovy, E.H. In *Proceedings of the 26th ACL Conference*, Buffalo, NY, June 1988. Also available as USC/Information Sciences Institute Research Report ISI/RS-88-209.

- On the Study of Text Planning and Realization. Hovy, E.H. In *Proceedings of AAAI Workshop on Text Planning and Realization*, St. Paul, MN, June 1988.

Acknowledgments

The work described in this document would have been impossible without the interest and comments of numerous people who helped to improve it, though all errors remain that of the author. It is impossible to thank them all adequately. The following people, however, deserve special mention:

For Section 3.1: Thanks for discussions to Dr. Cécile Paris (USC/ISI) and Dr. Nicholas Asher (University of Texas and University of Stuttgart).

For Section 3.3.2: Thanks to Dr. William Mann (retired from USC/ISI), Dr. Cécile Paris (USC/ISI), Prof. Johanna Moore (University of Pittsburgh), and Dr. William Swartout (USC/ISI).

For Section 3.3.5: Thanks to Prof. Kathleen McCoy (University of Delaware) and Dr. Julia Lavid (University of Madrid).

For Section 3.3.6: Thanks to Prof. Donia Scott (Brighton Polytechnic).

For Section 3.3.7: Thanks to Dr. Yigal Arens, and also to Dr. William Swartout for the idea of using LaTeX to illustrate the point, to Mr. Richard Whitney for implementation help, and to Dr. Lewis Johnson for ARIES (all from USC/ISI).

For Sections 3.4 and 3.3.3: Thanks to the ISI Text Planning Group: Dr. Cécile Paris (USC/ISI), Dr. Julia Lavid (University of Madrid), Ms. Elisabeth Maier (IPSI Institute, Darmstadt), Mr. Richard Whitney, USC/ISI, and Mr. Vibhu Mittal, USC, and to visitors Mr. Giuseppe Carenini (IRST, Trento) and Mr. Thanasis Daradoumis (University of Barcelona). Also thanks for discussions to Dr. Gisela Redeker, Ted Sanders, and Wilbert Spooren (all three from Tilburg University).

For Sections 3.5 and 3.6: Thanks to Dr. Yigal Arens (USC/ISI) and Ms. Mira Vossers (University of Nijmegen).

6 References

References

[Allen 91] Allen, J.F. 1991. Discourse structure in the TRAINS project. In *Proceedings of the DARPA Speech and Natural Language Workshop*, Morgan Kaufmann.

[Allen & Perrault 80] Allen, J.F. and Perrault, C.R. 1980. Analyzing intention in utterances. *Artificial Intelligence* 15 (143-178).

[APA 83] *Publication Manual of the American Psychological Association* (third edition). 1983. Washington: American Psychological Association.

[Arens et al. 88] Arens, Y., Miller, L., Shapiro, S.C. & Sondheimer, N.K. 1988. Automatic Construction of User-Interface Displays. In *Proceedings of the 7th AAAI Conference*, St. Paul, MN. Also appears as USC/Information Sciences Institute Research Report RR-88-218.

[Arens & Hovy 90a] Arens Y., and Hovy, E.H. 1990. How to Describe What? Towards a Theory of Modality Utilization. In Proceedings of the 12th Annual Conference of the Cognitive Science Society (487-494).

[Arens & Hovy 90b] Arens Y., and Hovy, E.H. 1990. Text Layout as a Problem of Modality Selection. In *Proceedings of the 5th Conference on Knowledge-Based Specification*, RADC Workshop, Syracuse, NY (87-94).

[Arens, Hovy, & Vossers 91] Arens Y., Hovy, E.H., and Vossers, M. 1991. Descrying the Knowledge Underlying the Processing of Multimedia Instruction Manuals. Submitted to the Cognitive Science Society Conference.

[Aristotle 54] Aristotle. 1954. The Rhetoric. In *The Rhetoric and the Poetics of Aristotle*, W. Rhys Roberts (trans). New York: Random House.

[ASA 89] ASA, Inc. 1989. *Airman's Information Manual.* Aviation Supplies and Academics.

[Asher 91] Asher, N. 1991. *Abstract Object, Semantics and Anaphora.* Unpublished manuscript, University of Texas at Austin (forthcoming).

[Austin 65] Austin, J.L. 1965. *How to do Things with Words.* Oxford: Oxford University Press.

[Biber 89] Biber, D. 9189. A Typology of English Texts. *Linguistics* 27 (3-43).

[Birnbaum et al. 80] Birnbaum, L.A., Flowers, M., & McGuire, R. 1980. Towards an AI Model of Argumentation. In *Proceedings of the 1st AAAI Conference*, Stanford (195-198).

[Björklund & Virtanen 89] Björklund, M. and Virtanen, T. 1989. Variation in Narrative Structure: A Simple Text vs. an Innovative Work of Art. Presented at the *16th International Systemics Congress*, Helsinki.

[Bretin 83] Bretin, J. 1983. *Semiology of Graphics.* (trans. by J. Berg). University of Wisconsin Press.

[Cawsey 90] Cawsey, A. 1990. Generating Communicative Discourse. In *Current Research in Natural Language Generation*, Dale, R., Mellish, C., and Zock, M. (eds), Boston: Academic Press (75-102). Also presented at the *Second European Workshop on Language Generation*, Edinburgh.

[Charniak & Shimony 90] Charniak, E. and Shimony, S.E. 1990. Probabilistic Semantics for Cost Based Abduction. In *Proceedings of the 8th AAAI Conference* (107-111).

[CMS 82] *The Chicago Manual of Style* (thirteenth edition). 1982. Chicago: University of Chicago Press.

[Cohen 83] Cohen, R. 1983. *A Computational Model for the Analysis of Arguments.* Technical Report CSRG-151, University of Toronto, Toronto.

[Cohen & Levesque 85] Cohen, P.R. and Levesque, H.J. 1985. Speech Acts and Rationality. In *Proceedings of the 23rd ACL Conference*, Chicago (49-59).

[Cohen & Levesque 90] Cohen, P.R. and Levesque, H.J. 1990. Rational Interaction as the Basis for Communication. In Cohen, P., Morgan, J., and Pollack, M. (eds) *Intentions in Communication.* Cambridge: MIT Press.

[Conklin & McDonald 82] Conklin, E.J. and McDonald, D.D. 1982. Salience: The Key to the Selection Problem in Natural Language Generation. In *Proceedings of the 20th ACL Conference*, Toronto (129-135).

[Cullingford 78] Cullingford, R. 1978. *Script Application: Computer Understanding of Newspaper Stories.* Ph.D. dissertation, Yale University.

[Dahlgren 88] Dahlgren, K. 1988. *Naive Semantics for Natural Language Understanding.* Boston: Kluwer Academic Press.

[Dale 88] Dale, R. 1988. Generating Referring Expressions in a Domain of Objects and Processes. Ph.D. dissertation, University of Edinburgh.

[Daneš 74] Daneš, F. 1974. Functional Sentence Perspective and the Organization of the Text. In F. Daneš (ed.) *Papers on Functional Sentence Perspective.* The Hague: Mouton.

[Davey 79] Davey, A. 1979. *Discourse Production.* Edinburgh: Edinburgh University Press.

[De Beaugrande 80] De Beaugrande, R. 1980. *Text, Discourse and Process.* Ablex Publishing Corporation.

[DeJong 79] DeJong, G.F. 1979. *Skimming Stories in Real Time: An Experiment in Integrated Understanding.* Ph.D. dissertation, Yale University.

[De Souza et al. 89] De Souza, C.S., Scott, D.R. and Nunes, M.G.V. 1989. Enhancing Text Quality in a Question-Answering System. Unpublished manuscript, Pontificia Universidade Católica de Rio de Janeiro.

[Dobeš & Novak 91] Dobeš, Z. and Novak, H-J. 1991. From Constituent Planning to Text Planning. In *Proceedings of the 3rd European Workshop on Language Generation*, Innsbruck, Austria (46-54).

[Dwyer 78] Dwyer, F.M. 1978. *Strategics for improving visual learning.* State College, PA: Learning Services.

[Dyer 83] Dyer, M.G. 1983. *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* Cambridge: MIT Press.

[Feiner 88] Feiner, S. 1988. An Architecture for Knowledge-Based Graphical Interfaces. In *Proceedings of the ACM/SIGCHI workshop on Architectures for Intelligent Interfaces: Elements and Prototypes*, Monterey, CA.

[Fleming & Levie 78] Fleming, M. and Levie, H.W. 1978. *Instructional Message Design: Principles from the Behavioral Sciences.* New Jersey: Educational Technology Publications.

[Fries 81] Fries, P.H. 1981. On the Status of Theme in English: Arguments from Discourse. *Forum Linguisticum* 6 (1-38).

[Giora 88] Giora, R. 1988. On the Informativeness Requirement. *Journal of Pragmatics* 12 (547-565).

[Grimes 75] Grimes, J.E. 1975. *The Thread of Discourse.* Mouton: The Hague.

[Grosz & Sidner 86] Grosz, B.J. and Sidner, C.L. 1986. Attention, Intentions, and the Structure of Discourse. In *Journal of Computational Linguistics* 12(3) (175-204).

[Grosz & Sidner 90] Grosz, B.J. and Sidner, C.L. 1990. Plans for Discourse. In Cohen, P., Morgan, J., and Pollack, M. (eds) *Intentions in Communication.* MIT Press.

[Halliday 67] Halliday, M.A.K. 1967. Notes on Transitivity and Theme in English. *Journal of Linguistics* 3 (37-81), 4 (179-215).

[Halliday 85] Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* Baltimore: Edward Arnold Press.

[Halliday & Hasan 76] Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English.* London: Longman.

[Halliday & Hasan 85] Halliday, M.A.K. and Hasan, R. 1985. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective.* Deakin University Press.

[Hartley 85] Hartley, J. 1985. *Designing Instructional Text.* (2nd edition). Great Britain: Kogan Page Ltd.

[Hobbs 78] Hobbs, J.R. 1978. Why is Discourse Coherent? Technical Note no. 176, SRI International, Menlo Park CA.

[Hobbs 79] Hobbs, J.R. 1979. Coherence and Coreference. *Cognitive Science* 3(1) (67-90).

[Hobbs 90] Hobbs, J.R. 1990. *Literature and Cognition.* CSLI Lecture Notes no. 21.

[Hobbs et al. 88] Hobbs, J.R., Stickel, M., Martin, P. & Edwards, D. 1988. Interpretation as Abduction. In *Proceedings of the 26th ACL Conference*, Buffalo, NY (95-103).

[Honda Manual] Honda Accord: 1990 Owner's Manual. Japan: Honda Motor Co. Ltd.

[Horacek 92] Horacek, H. 1992. An Integrative View of Text Planning. In *Proceedings of the Sixth International Workshop on Language Generation*, Trento, Italy, (to appear).

[Hovy 87] Hovy, E.H. 1987. Interpretation in Generation. In *Proceedings of the 6th AAAI Conference*, Seattle, WA (545-549). Also available as USC/Information Sciences Institute Research Report ISI/RS-88-186.

[Hovy 88] Hovy, E.H. 1988. Planning Coherent Multisentential Text. In *Proceedings of the 26th ACL Conference*, Buffalo, NY (163-169).

[Hovy 90a] Hovy, E.H. 1990. Approaches to the Planning of Coherent Text. In *Natural Language in Artificial Intelligence and Computational Linguistics*, Paris, C.L., Swartout, W.R. and Mann, W.C. (eds), Boston: Kluwer (83-102). Also presented at the *4th International Workshop on Text Generation*, Los Angeles, 1988.

[Hovy 90b] Hovy, E.H. 1990. Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. In *Proceedings of the 5th International Workshop on Text Generation*, Pittsburgh.

[Hovy 90c] Hovy, E.H. 1990. Natural Language Processing at ISI. In *The Finite String* 16(4) (37-42).

[Hovy & Arens 90] Hovy, E.H. and Arens, Y. 1990. When is a Picture Worth a Thousand Words? — Allocation of Modalities in Multimedia Communication. Presented at the *AAAI Symposium on Human-Computer Interaction*, Stanford University.

[Hovy & Arens 91] Hovy, E.H. and Arens, Y. 1991. Automatic Generation of Formatted Text. In *Proceedings of the 8th AAAI Conference*, Anaheim.

[Hovy & Lavid 92] Hovy, E.H. and Lavid, J.M. 1992. Focus, Theme, Given, and Other Dangerous Things. Working paper.

[Hovy & Maier 92] Hovy, E.H. and Maier, E. 1992. Parsimonious and Profligate: How Many and Which Discourse Structure Relations? Submitted to *Computational Intelligence*.

[Hovy & McCoy 89] Hovy, E.H. and McCoy, K.F. 1989. Focusing your RST: A step toward generating coherent multisentential text. In *Proceedings of the 11th Cognitive Science Conference*, Ann Arbor (667-674).

[Hovy et al. 92] Hovy, E.H., Lavid, J., Maier, E., Mittal, V., and Paris, C.L. 1992. Employing Knowledge Resources in a New Text Planning Architecture. In *Proceedings of the 6th International Workshop on Language Generation*, Trento, Italy (to appear).

[Ivir et al. 80] Ivir, V., McMillan, D. and Merz, T. 1980. S-Relators. Unpublished manuscript, University of Zagreb.

[Johnson & Harris 90] Johnson, W.L. and Harris, D. 1990. Requirements Analysis Using ARIES: Themes and Examples. In Proceedings of the 5th Knowledge Based Software Engineering Conference (79-84). Tech Report no. RL-TR-91-11, Rome Laboratory.

[Johnson & Feather 91] Johnson, W.L. and Feather, M. 1991. Using Evolution Transformations to Construct Specifications. In Lowry, P. and McCartney, J. eds. *Automating Software Design*, AAAI Press.

[Kamp 81] Kamp, H. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, Groenendijk, J.A.G., Janssen, T.M.V. and Stokhof, M.B.J. (eds), Mathematical Centre Tracts (vol 136), Amsterdam, (277-322).

[Kasper 89] Kasper, R.T. 1989. Unification and Classification: An Experiment in Information-Based Parsing. In *Proceedings of the International Workshop on Parsing Technologies*, Pittsburgh, PA.

[Kasper & Hovy 90] Kasper, R.T. and Hovy, E.H. 1990. Integrated Semantic and Syntactic Parsing using Classification. In *Proceedings of the Darpa Speech and Natural Language Workshop*, Pittsburgh, PA.

[Kautz 87] Kautz, H. 1987. *A Formal Theory of Plan Recognition.* Ph.D. dissertation, University of Rochester.

[Lambert & Carberry 91] Lambert, L. and Carberry, S. 1991. A Tripartite Plan-Based Model of Dialogue. In *Proceedings of the 29th Annual Conference of the Association of Computational Linguistics*, Berkeley (47-54).

[Lamport 86] Lamport, L. 1986. LaTeX *User's Guide and Reference Manual.* Reading: Addison-Wesley.

[Larkin & Simon 86] Larkin, J.H., and Simon, H.A. 1986. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11 (65-99).

[Lascarides & Asher 91] Lascarides, A. & Asher, N. 1991. Discourse Relations and Defeasible Knowledge. In *Proceedings of the 29th ACL Conference*, Berkeley.

[Leech 83] Leech, G.N. 1983. *Principles of Pragmatics.* New York and London: Longmans,

[Levy 84] Levy, E. 1984. *Communicating Thematic Structures in Narrative Discourse: The Use of Referring Terms and Gestures.* Ph.D. dissertation, University of Chicago.

[Litman 85] Litman, D. 1985. *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues.* Ph.D. dissertation, University of Rochester.

[Lochbaum 91] Lochbaum, K.E. 1991. An Algorithm for Plan Recognition in Collaborative Discourse. In *Proceedings of the 29th Annual Conference of the Association of Computational Linguistics*, Berkeley (33-38).

[MacGregor 88] MacGregor, R. 1988. A Deductive Pattern Matcher. In Proceedings of the 6th National Conference on Artificial Intelligence (AAAI-88), 696-701.

[Mackinlay 86] Mackinlay, J. 1986. *Automatic Design of Graphical Presentations.* Ph.D. dissertation, Stanford University.

[Maier & Hovy 91] Maier, E. and Hovy, E.H. 1991. A Metafunctionally Motivated Taxonomy for Discourse Structure Relations. In *Proceedings of the 3rd European Workshop on Language Generation*, Innsbruck, Austria (38-45).

[Mann 83] Mann, W.C. 1983. An Overview of the Penman Text Generation System. USC/Information Sciences Institute Research Report ISI/RR-83-114.

[Mann 84] Mann, W.C. 1984. Discourse Structures for Text Generation. In *Proceedings of the COLING Conference*, Stanford, CA.

[Mann 87] Mann, W.C. 1987. Text Generation: The Problem of Text Structure. USC/Information Sciences Institute Research Report RR-87-181.

[Mann & Matthiessen 83] Mann, W.C. and Matthiessen, C.M.I.M. 1983. Nigel: A Systemic Grammar for Text Generation. Research Report RR-83-105, USC/ISI.

[Mann & Matthiessen 85] Mann, W.C. and Matthiessen, C.M.I.M. 1985. Nigel: A Systemic Grammar for Text Generation. In Benson, R. and Greaves, J. (eds), *Systemic Perspectives on Discourse: Selected Papers Papers from the 9th International Systemics Workshop*. London: Ablex. Also available as USC/ISI Research Report RR-83-105.

[Mann & Moore 81] Mann, W.C. and Moore, J.A. 1981. Computer Generation of Multi-Paragraph English Text. *Computational Linguistics* 7(1) (63-89).

[Mann & Thompson 86] Mann, W.C. and Thompson, S.A. 1986. Rhetorical Structure Theory: Description and Construction of Text Structures. In *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics*, Kempen, G. (ed), Boston: Kluwer Academic Publishers (279-300).

[Mann & Thompson 88] Mann, W.C. and Thompson, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3) (243-281). Also available as USC/Information Sciences Institute Research Report RR-87-190.

[Marslen-Wilson et al. 82] Marslen-Wilson, W., Levy, E., and Tyler, L.K. 1982. Producing interpretable discourse: The establishment and maint of reference. In R.J. Jarvella and W. Klein (eds) *Speech, Place and Action*, New York: John Wiley and Sons (339-378).

[Martin 92] Martin, J.R. 1992. *English Text: System and Structure*. Amsterdam: Benjamins. In press.

[Matthiessen 84] Matthiessen, C.M.I.M. 1984. Systemic Grammar in Computation: The Nigel Case. In *Proceedings of 1st Conference of the European Association for Computational Linguistics*, Pisa, Italy. Also available as USC/ISI Research Report RR-84-121.

[Maybury 90] Maybury, M.T. 1990. *Planning Multisentential English Text Using Communicative Acts*. Ph.D. dissertation, Cambridge University. Also available as RADC Technical Report 90-411.

[Mayer 89] Mayer, R.E. 1989. Systematic Thinking Fostered by Illustrations in Scientific Text. *Journal of Educational Psychology* 81 (240-246).

[McCoy 85] McCoy, K.F. 1985. *Correcting Object-Related Misconceptions*. Ph.D. dissertation, University of Pennsylvania.

[McCoy & Cheng 88] McCoy, K.F. and Cheng, J. 1988. Focus of Attention: Constraining What can be Said Next. In *Natural Language in Artificial Intelligence and Computational Linguistics*, Paris, C.L., Swartout, W.R. and Mann, W.C. (eds), Boston: Kluwer (103-124). Presented at the *4th International Workshop on Text Generation*, Los Angeles.

[McKeown 85] McKeown, K.R. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge: Cambridge University Press.

[Meehan 76] Meehan, J. 1976. *The Metanovel: Writing Stories by Computer.* Ph.D. dissertation, Yale University.

[Meteer 90] Meteer, M.W. 1990. *The Generation Gap: The Problem of Expressibility in Text Planning.* Ph.D. Dissertation, University of Massachusetts at Amherst.

[Mooney et al. 90] Mooney, D.J., Carberry, S., and McCoy, K.F. 1990. The Basic Block Model of Extended Explanations. In *Proceedings of the 5th International Workshop on Text Generation*, Pittsburgh.

[Moore 89] Moore, J.D. 1989. *A Reactive Approach to Explanation in Expert and Advice-Giving Systems.* Ph.D. dissertation, University of California in Los Angeles.

[Moore & Paris 89] Moore, J.D. and Paris, C.L. 1989. Planning Text for Advisory Dialogues. In *Proceedings of the 27th ACL Conference*, Vancouver (67-75).

[Moore & Paris 91] Moore, J.D. and Paris, C.L. 1991. Discourse Structure for Explanatory Dialogues. Presented at the AAAI Fall Symposium on Discourse, Pacific Grove.

[Moore & Swartout 90] Moore, J.D. and Swartout, W.R. 1990. Dialogue-Based Explanation. In *Natural Language in Artificial Intelligence and Computational Linguistics*, Paris, C.L., Swartout, W.R. and Mann, W.C. (eds), Boston: Kluwer (3-48). Also presented at the *4th International Workshop on Text Generation*, Los Angeles, 1988.

[Morris & Hirst 91] Morris, J. and Hirst, G. 1991. Lexical Cohesion Computed by Theoretical Relations as an Indicator of the Structure of the Text. *Computational Linguistics* 11(1) (21-48).

[Neal 90] Neal, J.G. 1990. Intelligent Multi-Media Integrated Interface Project. SUNY Buffalo. RADC Technical Report TR-90-128.

[Paris 87] Paris, C.L. 1987. The Use of Explicit Models in Text Generation: Tailoring to a User's Level of Expertise. Ph.D. dissertation, Columbia University.

[Paris 90] Paris, C.L. 1990. Generation and Explanation: Building an Explanation Facility for the Explainable Expert Systems Framework. In *Natural Language in Artificial Intelligence and*

Computational Linguistics, Paris, C.L., Swartout, W.R. and Mann, W.C. (eds), Boston: Kluwer (49-82). Also presented at the *4th International Workshop on Text Generation*, Los Angeles, 1988.

[Paris 91] Paris, C.L. 1991. The Role of the User's Domain Knowledge in Generation. *Computational Intelligence* 7(2) (71-93).

[Passoneau 91] Passoneau, R.J. 1991. Getting and keeping the center of attention. In R. Weischedel and M. Bates (eds) *Challenges in Natural Language Processing*, Cambridge: Cambridge University Press.

[Penman 89] *The Penman Primer, User Guide, and Reference Manual, and the Nigel Manual.* Unpublished Penman documentation, USC/ISI, 1988.

[Petre & Green 90] Petre, M. and Green, T.R.G. 1990. Is Graphical Notation Really Superior to Text, or Just Different? Some Claims by Logic Designers about Graphics in Notation. In *Proceedings of the ECCE-5*, Urbino, Italy.

[Polanyi 88] Polanyi, L. 1988. A formal Model of the Structure of Discourse. *Journal of Pragmatics* 12 (601-638).

[Pollack 86] Pollack, M.E. 1986. A Model of Plan Inference that Distinguishes between the Beliefs of Actors and Observers. In *Proceedings of the 24th Annual Conference of the Association of Computational Linguistics*, Chicago (207-214).

[Quirk et al. 72] Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1972. *A Grammar of Contemporary English.* New York: Seminar Press.

[Quirk & Greenbaum 73] Quirk, R., and Greenbaum, S. 1973. *A Concise Grammar of Contemporary English.* New York: Harcourt Brace Jovanovich Inc.

[Rambow 90] Rambow, O. 1990. Domain Communication Knowledge. In *Proceedings of the 5th International Workshop on Text Generation*, Pittsburgh (87-94).

[Rambow & Korelsky 92] Rambow, O. and Korelsky, T. 1992. Applied Text Generation. In *Proceedings of the Applied Natural Language Processing Conference*, Trento (to appear).

[Ramshaw 91] Ramshaw, L.A. 1991. A Three-Level Model for Plan Exploration. In *Proceedings of the 29th Annual Conference of the Association of Computational Linguistics*, Berkeley (39-46).

[Rankin 89] Rankin, I. 1989. *The Deep Generation of Text in Expert Critiquing Systems.* Licentiate thesis, University of Linköping, Sweden.

[Redeker 90] Redeker, G. 1990. Ideational and Pragmatic Markers of Discourse Structure. *Journal of Pragmatics* 14 (367-381).

[Reichman 85] Reichman, R. 1985. *Getting Computers to Talk Like You and Me.* Cambridge: MIT Press.

[Reithinger 91] Reithinger, N. 1991. *Eine Parallele Architektur zur Inkrementellen Generierung Multimodaler Dialogbeiträge*, Ph.D. dissertation, University of the Saarland, Germany.

[Roth & Mattis 90] Roth, S.F. and Mattis, J. 1990. Data Characterization for Intelligent Graphics Presentation. In *Proceedings of the CHI Conference* (193-200).

[Rumelhart 72] Rumelhart, D.E. 1972. Notes on a Schema for Stories. In Bobrow, D.G. and Collins, A. (eds) *Representation and Understanding.* New York: Academic Press.

[Sacerdoti 77] Sacerdoti, E. 1977. *A Structure for Plans and Behavior.* Amsterdam: North-Holland.

[Sanders et al. 91] Sanders, T.J.M., Spooren, W.P.M.S., and Noordman, L.G.M. 1991. Towards a Taxonomy of Coherence Relations. In *Discourse Processes* (forthcoming).

[Schank & Abelson 77] Schank, R.C. and Abelson, R. 1977. *Scripts, Plans, Goals, and Understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[Scott & De Souza 90] Scott, D.R. and De Souza, C.S. 1990. Getting the Message across in RST-Based Text Generation. In *Current Research in Natural Language Generation*, Dale, R., Mellish, C., and Zock, M. (eds), Bosont: Academic Press (47-74).

[Shepherd 26] Shepherd, H.R. 1926. *The Fine Art of Writing.* New York: The Macmillan Co.

[Sidner 83] Sidner, C.L. 1983. Focusing and Discourse. *Discourse Processes* 6 (107-130).

[Sycara 87] Sycara, K. 1987. *Resolving Adversarial Conflicts: An Approach Integrating Case-Based and Analytical Methods.* Ph.D. dissertation, Georgia Institute of Technology.

[Toulmin 58] Toulmin, S. 1958. *The Uses of Argument.* Cambridge: Cambridge University Press.

[Trimble 85] Trimble, L. 1985. *English for Science and Technology: A Discourse Approach.* Cambridge: Cambridge University Press.

[Tufte 83] Tufte, E.R. 1983. *The Visual Display of Quantitative Information.* Cheshire: Graphics Press.

[Tufte 90] Tufte, E.R. 1990. *Envisioning Information.* Cheshire: Graphics Press.

[Twyman 85] Twyman, M. 1985. Using Pictorial Language: A Discussion of the Dimensions of the Problem. In Duffy, T.M., and Waller, R. (eds.) *Designing Usable Texts.* Florida: Academic Press (245-312).

[Van Dijk 72] Van Dijk, T.A. 1972. *Some Aspects of Text Grammars.* The Hague: Mouton.

[Van Dijk & Kintsch 83] Van Dijk, T.A. and Kintsch, W. 1983. *Strategies of Discourse Comprehension.* New York: Academic Press.

[Van Leunen 79] Van Leunen, M.-C. 1979. *A Handbook for Scholars.* New York: Alfred A. Knopf.

[Ventola 87] Ventola, E. 1987. *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters.* London: Francis Pinter.

[Vossers 91] Vossers, M. 1991. Automatic Generation of Formatted text and Line Drawings. Master's thesis, University of Nijmegen, The Netherlands.

[Wahlster et al. 91] Wahlster, W., André, E., Bandyopadhyay, S., Graf, W., Rist, T. 1991. WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation. DFKI (University of Saarbrücken) Technical Report.

[Werth 84] Werth, P. 1984. *Focus, Coherence and Emphasis.* London: Croom Helm.

*U.S. GOVERNMENT PRINTING OFFICE: 1993-710-093-60146

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.

END

FILMED

DATE: 4-93

DTIC